

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 11:51:11 ; Search time 213.2 Seconds

(without alignments)
16.487 Million cell updates/sec

Title: US-10-698-121a-1

Perfect score: 42

Sequence: 1 GAGSPGL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

A Geneseq 21.*

1: Geneseqp1980s.*

2: Geneseqp1980s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	8	AD015278	Ad015278 Collagen
2	42	100.0	12	AD015279	Ad015279 Collagen
3	42	100.0	652	ABM83520	ABM83520 Human dia
4	42	100.0	654	ABM83519	ABM83519 Human dia
5	42	100.0	657	ABM83518	ABM83518 Human dia
6	42	100.0	661	ABM83517	ABM83517 Human dia
7	42	100.0	673	ABM83516	ABM83516 Human dia
8	42	100.0	683	ABM83515	ABM83515 Human dia
9	42	100.0	685	ABM83514	ABM83514 Human dia
10	42	100.0	688	ABM83529	ABM83529 Human dia
11	42	100.0	692	ABM83528	ABM83528 Human dia
12	42	100.0	695	ABM83527	ABM83527 Human dia
13	42	100.0	697	ABM83526	ABM83526 Human dia
14	42	100.0	699	ABM83525	ABM83525 Human dia
15	42	100.0	700	ABM83524	ABM83524 Human dia
16	42	100.0	702	ABM83523	ABM83523 Human dia
17	42	100.0	702	ABM83522	ABM83522 Human dia
18	42	100.0	705	ABR40108	ABR40108 Human cel
19	42	100.0	714	ABM83521	ABM83521 Human dia
20	42	100.0	717	ADN95515	ADN95515 Human BEC
21	39	92.9	495	ABO83203	ABO83203 Pseudomon
22	39	92.9	738	ADW44464	ADW44464 Murine co
23	38	90.5	30	AY233942	AY233942 Amino aci
24	38	90.5	30	AY233938	AY233938 Amino aci

25	38	90.5	134	4	ABU53024	Abu53024 Human tes
26	38	90.5	136	2	ABU53020	Abu53020 Human tes
27	38	90.5	168	2	AY23939	AY23939 Amino aci
28	38	90.5	219	3	AY84402	AY84402 C-termina
29	38	90.5	219	3	AY84555	AY84555 A C-termi
30	38	90.5	219	3	AY84553	AY84553 A C-termi
31	38	90.5	238	7	ABO67957	ABO67957 Pseudomon
32	38	90.5	317	9	ABE27174	ABE27174 Pinus rad
33	38	90.5	465	9	ADW87049	ADW87049 Human pan
34	38	90.5	498	9	ADW99569	ADW99569 Human gel
35	38	90.5	498	9	ADW99568	ADW99568 Human gel
36	38	90.5	516	7	ADE87046	ADE87046 Human pan
37	38	90.5	660	9	ADW99570	ADW99570 Human gel
38	38	90.5	660	9	ADW99571	ADW99571 Human gel
39	38	90.5	773	8	ADR08580	ADR08580 Human pro
40	38	90.5	886	7	ABO87052	ABO87052 Human pan
41	38	90.5	926	7	ABO83805	ABO83805 Pseudomon
42	38	90.5	1014	9	ADW99575	ADW99575 Human gel
43	38	90.5	1014	9	ADW99579	ADW99579 Human gel
44	38	90.5	1014	9	ADW99574	ADW99574 Human gel
45	38	90.5	1014	9	ADW99577	ADW99577 Human gel
46	38	90.5	1014	9	ADW99573	ADW99573 Human gel
47	38	90.5	1014	9	ADW99572	ADW99572 Human gel
48	38	90.5	1014	9	ADW99576	ADW99576 Human gel
49	38	90.5	1014	9	ADW99578	ADW99578 Human gel
50	38	90.5	1014	9	ADW99580	ADW99580 Human gel
51	38	90.5	1057	3	AY84544	AY84544 A human c
52	38	90.5	1058	3	AY84403	AY84403 Amino aci
53	38	90.5	1107	3	AY84540	AY84540 Amino aci
54	38	90.5	1161	7	ADE87050	ADE87050 Human pan
55	38	90.5	1171	3	AY84538	AY84538 A chimeri
56	38	90.5	1211	7	ADE87057	ADE87057 Human pan
57	38	90.5	1226	7	ADE87062	ADE87062 Human pan
58	38	90.5	1284	8	ADS98184	ADS98184 Protein f
59	38	90.5	1341	2	NAR71701	Nar71701 Collagen
60	38	90.5	1341	3	AY96122	AY96122 Collagen
61	38	90.5	1341	5	AAE16475	AAE16475 Human col
62	38	90.5	1341	5	ABB80733	ABB80733 Collagen
63	38	90.5	1341	5	ABB09625	ABB09625 Amino aci
64	38	90.5	1341	7	ADF13075	ADF13075 Human col
65	38	90.5	1388	3	AY84539	AY84539 Amino aci
66	38	90.5	1411	3	AY56800	AY56800 Human pre
67	38	90.5	1449	4	AAE02535	AAE02535 Porcine a
68	38	90.5	1453	5	ABG93948	ABG93948 Mouse pol
69	38	90.5	1453	7	ADD45053	ADD45053 Rat Prote
70	38	90.5	1453	7	ADD45057	ADD45057 Rat Prote
71	38	90.5	1453	7	ADD48341	ADD48341 Rat Prote
72	38	90.5	1453	7	ADD45049	ADD45049 Rat Prote
73	38	90.5	1453	7	ADD48337	ADD48337 Rat Prote
74	38	90.5	1453	7	ADD48345	ADD48345 Rat Prote
75	38	90.5	1453	9	ADW44458	ADW44458 Murine FV
76	38	90.5	1461	5	ABG93947	ABG93947 Human pol
77	38	90.5	1463	4	AAE02532	AAE02532 Bovine al
78	38	90.5	1464	2	AAW68485	AAW68485 Human rec
79	38	90.5	1464	4	RAU14136	RAU14136 Human nov
80	38	90.5	1464	7	ADD45059	ADD45059 Human pro
81	38	90.5	1464	7	ADD45055	ADD45055 Human pro
82	38	90.5	1464	7	ADD45051	ADD45051 Human pro
83	38	90.5	1464	7	ADE87048	ADE87048 Human pan
84	38	90.5	1464	8	ADR16800	ADR16800 Human col
85	38	90.5	1464	8	ADR16425	ADR16425 Human col
86	38	90.5	1464	9	ADR99144	ADR99144 Collagen,
87	38	90.5	1464	9	ADX58061	ADX58061 Amino aci
88	38	90.5	1518	4	ABG22679	ABG22679 Novel hum
89	38	90.5	1518	7	ADS98730	ADS98730 Protein f
90	38	90.5	1536	8	ADY24225	ADY24225 Plant ful
91	37	88.1	359	8	ADY24226	ADY24226 Plant ful
92	37	88.1	387	8	ADY13803	ADY13803 Plant ful
93	37	88.1	467	8	ADL68215	ADL68215 Plant ful
94	37	88.1	1027	8	ADL92144	ADL92144 Collagen
95	37	88.1	1690	4	AAW23916	AAW23916 Human EST
96	37	88.1	49	3	AAW23916	AAW23916 Human EST
97	36	85.7			AAW23916	AAW23916 Gene 1 hu

98	36	85.7	93	3	AAB34143	Aab34143 Gene 1 hu	171	35	83.3	260	6	ADA89693	Ada89693 Staphyloc
99	36	85.7	133	5	ABP05501	Abp05501 Human ORF	172	35	83.3	261	6	ADA89489	Ada89489 Staphyloc
100	36	85.7	176	5	ADK36502	Adk36502 Novel hum	173	35	83.3	261	6	ABM72140	Abm72140 Staphyloc
101	36	85.7	218	7	ADL25758	Adl25758 Human can	174	35	83.3	265	9	ADZ22859	Adz22859 INSP16-B
102	36	85.7	298	4	AUA48214	Aua48214 Propionib	175	35	83.3	271	9	ADZ22869	Adz22869 His-tagge
103	36	85.7	298	6	ABM44733	Abm44733 Propionib	176	35	83.3	293	6	ABU70813	Abu70813 Human adi
104	36	85.7	344	9	ADY64806	Ady64806 S. manson	177	35	83.3	294	6	ADZ22855	Adz22855 INSP161 m
105	36	85.7	376	9	ABM91655	Abm91655 M. xanthu	178	35	83.3	296	7	ADJ70976	Adj70976 Human hea
106	36	85.7	416	8	ABO59496	Abos9496 Human gen	179	35	83.3	300	9	ADZ22865	Adz22865 INSP161 H
107	36	85.7	484	2	AAB89274	Aaw89274 Granulocy	180	35	83.3	309	4	AUA23675	Aua23675 Novel hum
108	36	85.7	486	7	ADD47729	Add47729 Human Pro	181	35	83.3	309	4	ABG60248	Abg60248 Human ova
109	36	85.7	544	8	ADM48393	Adm48393 Recombina	182	35	83.3	309	5	ABG61719	Abg61719 Novel ova
110	36	85.7	619	2	AAM89271	Aam89271 Granulocy	183	35	83.3	339	6	ABU70735	Abu70735 Human adi
111	36	85.7	749	9	AEB91438	Aeb91438 Microbial	184	35	83.3	352	9	ADY65014	Ady65014 S. manson
112	36	85.7	994	7	ABM85518	Abm85518 Mouse pro	185	35	83.3	357	2	AAR95115	Aar95115 Interveni
113	36	85.7	1056	9	AEC04787	Aec04787 Human bre	186	35	83.3	357	2	AAM57646	Aaw57646 Collagen-
114	36	85.7	1081	9	AEC04183	Aec04183 Human bre	187	35	83.3	357	4	AAB64008	Aab64008 CLP-CB pr
115	36	85.7	1519	3	AAAY85660	Aay85660 Human tyr	188	35	83.3	357	4	AAB72738	Aab72738 Repetitiv
116	36	85.7	1603	8	ADQ21405	Adq21405 Human sof	189	35	83.3	377	7	ABO71733	Abot71733 Pseudomon
117	36	85.7	1745	4	AAG77793	Aag77793 Human pro	190	35	83.3	410	9	ADZ22857	Adz22857 INSP16-A
118	36	85.7	1745	5	ABB97234	Abb97234 Novel hum	191	35	83.3	416	4	AAB68065	Aab68065 Amino aci
119	36	85.7	1745	8	ADQ19841	Adq19841 Human sof	192	35	83.3	416	4	AAE02711	Aae02711 Human alp
120	36	85.7	1767	8	ADQ39813	Adq39813 Human myo	193	35	83.3	416	7	ADB84298	Adb84298 Recombina
121	36	85.7	1767	8	ADQ39817	Adq39817 Human myo	194	35	83.3	416	9	ADZ22867	Adz22867 His-tagge
122	36	85.7	1806	5	AAU84266	Aau84266 Human end	195	35	83.3	428	8	ADS30088	Ads30088 Bacterial
123	36	85.7	1806	5	ABJ05596	Abj05596 Breast ca	196	35	83.3	430	9	ABM94049	Abm94049 M. xanthu
124	36	85.7	1806	6	ABR58545	Abr58545 Human can	197	35	83.3	438	9	ADV86405	Adv86405 Novel Clq
125	36	85.7	1806	6	ABU56581	Abu56581 Lung canc	198	35	83.3	459	7	ADJ70748	Adj70748 Human hea
126	36	85.7	1806	7	ADP65251	Adp65251 Human alp	199	35	83.3	459	9	ADV86382	Adv86382 Novel Clq
127	36	85.7	1806	8	ADQ39816	Adq39816 Human myo	200	35	83.3	470	9	ADZ22875	Adz22875 INSP161 s
128	36	85.7	1806	8	ADQ39815	Adq39815 Human myo	201	35	83.3	476	9	ADZ22877	Adz22877 His-tagge
129	36	85.7	1806	9	ADZ09764	Adz09764 Human bre	202	35	83.3	477	9	ADV86418	Adv86418 Novel Clq
130	36	85.7	1806	9	AEC04181	Aec04181 Human bre	203	35	83.3	500	4	AAB68062	Aab68062 Amino aci
131	36	85.7	1806	9	AEC04182	Aec04182 Human bre	204	35	83.3	500	7	ADB84295	Adb84295 Recombina
132	36	85.7	1818	8	ADQ39814	Adq39814 Human myo	205	35	83.3	510	4	AAB68066	Aab68066 Amino aci
133	36	85.7	1818	8	ADQ39813	Adq39813 Human myo	206	35	83.3	510	4	AAE02712	Aae02712 Recombina
134	35	83.3	62	2	AAR05303	Aar05303 Collagen	207	35	83.3	510	4	AAE02712	Aae02712 Recombina
135	35	83.3	62	2	AAW07537	Aaw07537 Interveni	208	35	83.3	510	7	ADB84299	Adb84299 Recombina
136	35	83.3	62	4	AAB64005	Aab64005 CLP-CB mo	209	35	83.3	510	7	AAW06239	Aay06239 Mouse rec
137	35	83.3	62	4	AAB72735	Aab72735 Repetitiv	210	35	83.3	623	2	AAW12843	Aaw12843 Pro-alpha
138	35	83.3	63	5	AAU75471	Aau75471 Collagen-	211	35	83.3	626	2	AAW12842	Aaw12842 Truncated
139	35	83.3	63	6	ABR43719	Abr43719 Collagen	212	35	83.3	656	8	ADN25129	Adn25129 Bacterial
140	35	83.3	63	6	ADJ55711	Adj55711 Collagen	213	35	83.3	662	4	AAB68072	Aab68072 Amino aci
141	35	83.3	63	9	ADZ58917	Adz58917 Protein p	214	35	83.3	662	4	AAE02718	Aae02718 Human alp
142	35	83.3	63	9	ABE11736	Aeb11736 Collagen-	215	35	83.3	662	7	ADB84306	Adb84306 Recombina
143	35	83.3	63	9	ABE47080	Aeb47080 Engineere	216	35	83.3	770	7	ADG10784	Adg10784 Human STA
144	35	83.3	69	2	AAR95146	Aar95146 Collagen	217	35	83.3	770	7	ADG10782	Adg10782 Human STA
145	35	83.3	69	4	AAB64041	Aab64041 CLP/CB-F	218	35	83.3	822	2	AAW06240	Aay06240 Mouse rec
146	35	83.3	69	4	AAB72771	Aab72771 Repetitiv	219	35	83.3	822	8	ADY98120	Ady98120 Protein f
147	35	83.3	72	2	AAB95148	Aar95148 Collagen	220	35	83.3	825	6	ABP71596	Abp71596 HSV-2 ICP
148	35	83.3	72	4	AAB64043	Aab64043 CLP/CB-L2	221	35	83.3	825	7	ADG75089	Adg75089 Human her
149	35	83.3	72	4	AAB72773	Aab72773 Repetitiv	222	35	83.3	826	5	AAE17827	Aae17827 Herpes si
150	35	83.3	82	2	AAR95149	Aar95149 Collagen	223	35	83.3	826	7	ADG74975	Adg74975 Human her
151	35	83.3	85	4	AAB64044	Aab64044 CLP/CB-F	224	35	83.3	832	7	ADF74153	Adf74153 Human nov
152	35	83.3	85	4	AAB72774	Aab72774 Repetitiv	225	35	83.3	832	7	ADJ69970	Adj69970 Human hea
153	35	83.3	97	4	ABE64153	Abb64153 Drosophil	226	35	83.3	921	9	ADY95810	Ady95810 Murine ty
154	35	83.3	112	2	AAR05304	Aar05304 Collagen	227	35	83.3	966	7	ADG10666	Adg10666 Human STA
155	35	83.3	115	9	ABM91684	Abm91684 M. xanthu	228	35	83.3	966	7	ADG10668	Adg10668 Human STA
156	35	83.3	167	4	AAB68064	Aab68064 Amino aci	229	35	83.3	966	7	ADG10670	Adg10670 Human STA
157	35	83.3	167	4	AAE02710	Aae02710 Human alp	230	35	83.3	966	7	ADG10786	Adg10786 Human STA
158	35	83.3	167	7	ADB84297	Adb84297 Recombina	231	35	83.3	967	7	ABM85519	Abm85519 Human pro
159	35	83.3	197	4	AAU17725	Aau17725 Novel hum	232	35	83.3	1028	8	ADN35278	Adn35278 Helical d
160	35	83.3	197	7	ADG41105	Adg41105 Human res	233	35	83.3	1040	9	ADZ70363	Adz70363 Human pro
161	35	83.3	197	7	ADI96879	Adi96879 Human res	234	35	83.3	1057	3	AAAY84541	Aay84541 Amino aci
162	35	83.3	240	4	ABG08371	Abg08371 Novel hum	235	35	83.3	1078	3	AAW71704	Aaw71704 Collagen
163	35	83.3	250	8	ADS91505	Ads91505 Partial h	236	35	83.3	1078	3	AAAY96125	Aay96125 Collagen
164	35	83.3	251	4	AAB68071	Aab68071 Amino aci	237	35	83.3	1078	5	AAE16478	Aae16478 Human col
165	35	83.3	251	4	AAB68061	Aab68061 Amino aci	238	35	83.3	1078	5	ABE80736	Abe80736 Collagen
166	35	83.3	251	4	AAE02707	Aae02707 Human alp	239	35	83.3	1078	5	ABO96828	Abbo96828 Amino aci
167	35	83.3	251	4	AAE02717	Aae02717 Human alp	240	35	83.3	1078	7	ADF13078	Adf13078 Human col
168	35	83.3	251	7	ADB84304	Adb84304 Recombina	241	35	83.3	1107	2	AAR89472	Aar89472 Collagen/
169	35	83.3	251	7	ADB84294	Adb84294 Recombina	242	35	83.3	1169	2	AAR89469	Aar89469 Collagen/
170	35	83.3	252	8	ADS91506	Ads91506 Partial R	243	35	83.3	1169	3	AAAY84537	Aay84537 Amino aci

244	35	83.3	1171	2	AAR89470	Aar89470 Collagen/	317	34	81.0	153	8	ADL31452	Adl31452 Human pro
245	35	83.3	1196	2	AAR28916	Abu28916 Type III	318	34	81.0	154	4	AAO00600	Aao00600 Human pol
246	35	83.3	1306	6	ABU36481	Abu36481 Protein e	319	34	81.0	163	6	ABR48481	Abr48481 Human Alp
247	35	83.3	1306	9	ABE91429	Aeb91429 Microbial	320	34	81.0	164	6	AAU52505	Aau52505 Propionib
248	35	83.3	1313	8	ADN35279	Adn35279 Synthetic	321	34	81.0	164	6	ABM49024	Abm49024 Propionib
249	35	83.3	1313	8	ADN35277	Adn35277 Helical d	322	34	81.0	173	4	ABG65925	Abg65925 Drosophila
250	35	83.3	1388	2	AAR89471	Aar89471 Collagen/	323	34	81.0	176	7	ABO72106	Abu72106 Pseudomon
251	35	83.3	1424	8	ADQ14315	Adq14315 Human col	324	34	81.0	209	7	ADM05685	Adm05685 Human pro
252	35	83.3	1424	8	ADR41697	Adr41697 Human col	325	34	81.0	243	4	ABG95566	Abg95566 Human pro
253	35	83.3	1424	8	ADU06693	Adu06693 Novel bro	326	34	81.0	244	4	AAU32672	Aau32672 Novel hum
254	35	83.3	1424	9	ABE77777	Aeb77777 Human goo	327	34	81.0	264	7	ADD48444	Add48444 Human pol
255	35	83.3	1464	4	ABE82454	AbE82454 Human pro	328	34	81.0	284	7	ADF58956	Adf58956 Human pol
256	35	83.3	1464	5	ABB90764	Abb90764 Human tum	329	34	81.0	293	3	AAV52424	Aav52424 Human gly
257	35	83.3	1464	5	ABP68610	Abp68610 Human pan	330	34	81.0	293	5	AAO18245	Aao18245 Cell gly
258	35	83.3	1464	6	ABU54471	Abu54471 Human tum	331	34	81.0	293	5	AAO18247	Aao18247 Cell gly
259	35	83.3	1464	6	ABR47417	Abr47417 Breast ca	332	34	81.0	295	3	AAV52423	Aav52423 Human gly
260	35	83.3	1464	6	ABR92064	AbR92064 Human cer	333	34	81.0	295	3	AAO18246	Aao18246 Cell gly
261	35	83.3	1464	7	ADD14142	Add14142 Human src	334	34	81.0	295	5	AAO18244	Aao18244 Cell gly
262	35	83.3	1464	7	ADP65246	Adp65246 Human alp	335	34	81.0	295	5	ADQ88272	Adq88272 Human 164
263	35	83.3	1464	8	ADQ15470	Adq15470 Human sof	336	34	81.0	295	8	ADQ88272	Adq88272 Human 164
264	35	83.3	1464	8	ADQ29653	Adq29653 Human col	337	34	81.0	307	4	AAU45713	Aau45713 Propionib
265	35	83.3	1464	9	ADV87000	Adv87000 Collagen	338	34	81.0	307	6	ABM42232	Abm42232 Propionib
266	35	83.3	1464	9	ADV70233	Adv70233 Tumor-ass	339	34	81.0	355	7	ABO78680	Abu78680 Pseudomon
267	35	83.3	1464	9	ADZ26573	Adz26573 Human typ	340	34	81.0	379	8	ADS44333	Ads44333 Bacterial
268	35	83.3	1464	9	ADZ09874	Adz09874 Human bre	341	34	81.0	403	3	ABG17301	Abg17301 Novel hum
269	35	83.3	1464	9	AEA04480	Aea04480 Human pro	342	34	81.0	419	4	ABM82895	Abm82895 Human dia
270	35	83.3	1466	4	AAE02537	Aae02537 Porcine a	343	34	81.0	433	8	ABM82895	Abm82895 Human ptp
271	35	83.3	1466	4	AAE02534	Aae02534 Bovine al	344	34	81.0	443	9	ADR73494	Adr73494 Human lip
272	35	83.3	1466	4	AAE02533	Aae02533 Bovine al	345	34	81.0	469	7	ADE61178	Ade61178 Rat Prote
273	35	83.3	1466	4	ABB50291	Abb50291 Collagen	346	34	81.0	482	9	ADV86415	Adv86415 Novel Clq
274	35	83.3	1466	5	ABB90747	Abb90747 Human tum	347	34	81.0	495	2	AAW39747	Aaw39747 Human mar
275	35	83.3	1466	5	ABU54454	Abu54454 Human tum	348	34	81.0	520	2	AAW39748	Aaw39748 Human mar
276	35	83.3	1466	6	ABR47418	Abr47418 Breast ca	349	34	81.0	520	2	AAW39748	Aaw39748 Human mar
277	35	83.3	1466	7	ADP65248	Adp65248 Human alp	350	34	81.0	520	2	AAW39748	Aaw39748 Human mar
278	35	83.3	1466	7	ADP65210	Adp65210 Human alp	351	34	81.0	520	4	AAU29189	Aau29189 Human PRO
279	35	83.3	1466	8	ADQ26091	Adq26091 Type III	352	34	81.0	520	4	AAU29189	Aau29189 Human PRO
280	35	83.3	1466	8	ADQ29677	Adq29677 Human col	353	34	81.0	520	5	ABB83159	Abb83159 Human MAR
281	35	83.3	1466	8	ADR16802	Adr16802 Human col	354	34	81.0	520	6	ABU88113	Abu88113 Novel hum
282	35	83.3	1466	8	ADR16427	Adr16427 Human col	355	34	81.0	520	6	ABU88113	Abu88113 Novel hum
283	35	83.3	1466	8	ABM80366	Abm80366 Tumor-ass	356	34	81.0	520	6	ABU84428	Abu84428 Human sec
284	35	83.3	1466	8	ADR67267	Adr67267 Human bla	357	34	81.0	520	6	ABR66302	Abr66302 Human sec
285	35	83.3	1466	9	ADV70235	Adv70235 Tumor-ass	358	34	81.0	520	6	ABU99632	Abu99632 Human sec
286	35	83.3	1466	9	ADZ09873	Adz09873 Human bre	359	34	81.0	520	6	ABU99632	Abu99632 Human sec
287	35	83.3	1466	9	AEA04495	Aea04495 Human pro	360	34	81.0	520	6	ABU89992	Abu89992 Novel hum
288	35	83.3	1469	4	ABG15191	Abg15191 Novel hum	361	34	81.0	520	6	ABR68241	Abr68241 Human sec
289	35	83.3	1470	8	ADU04510	Adu04510 SLPI-Coll	362	34	81.0	520	6	ABR68241	Abr68241 Human sec
290	35	83.3	1572	8	ADU04510	Adu04510 Modified	363	34	81.0	520	6	ABU96294	Abu96294 Novel hum
291	35	83.3	1604	8	ADM84483	Adm84483 Human dia	364	34	81.0	520	6	ABU96294	Abu96294 Novel hum
292	35	83.3	1611	8	ABM84483	Abm84483 Human dia	365	34	81.0	520	6	ABU92725	Abu92725 Human sec
293	35	83.3	1642	8	AEA64562	Aea64562 Human lup	366	34	81.0	520	6	ABO08802	Abu08802 Human sec
294	35	83.3	1670	7	ADD47063	Add47063 Human pro	367	34	81.0	520	6	ABO08802	Abu08802 Human sec
295	35	83.3	1670	8	AEA64561	Aea64561 Human lup	368	34	81.0	520	6	ABO08802	Abu08802 Human sec
296	35	83.3	1676	3	AAV92060	Aav92060 Murine AP	369	34	81.0	520	6	ABR94770	Abr94770 Human sec
297	35	83.3	1726	6	ABR42661	Abr42661 Decorin-m	370	34	81.0	520	6	ABO25278	Abu25278 Novel hum
298	35	83.3	1950	8	ADU04493	Adu04493 Modified	371	34	81.0	520	6	ABU85743	Abu85743 Human sec
299	35	83.3	2274	4	ABM50674	Abm50674 Mouse APC	372	34	81.0	520	6	ABU98903	Abu98903 Novel hum
300	35	83.3	2274	8	ADU08047	Adu08047 Mouse pol	373	34	81.0	520	6	ABU98903	Abu98903 Novel hum
301	35	83.3	2274	8	ADH5091	Adh5091 Full-leng	374	34	81.0	520	6	ABU91824	Abu91824 Novel hum
302	35	83.3	3295	8	ADH39702	Adh39702 Streptomy	375	34	81.0	520	6	ABU91824	Abu91824 Novel hum
303	35	83.3	3651	8	ADH39704	Adh39704 Streptomy	376	34	81.0	520	6	ABU67571	Abu67571 Human sec
304	34	81.0	9	2	AAW49155	Aaw49155 Human leu	377	34	81.0	520	6	ABU67571	Abu67571 Human sec
305	34	81.0	50	5	ABP10034	Abp10034 Human ORF	378	34	81.0	520	6	ABR9517	Abr9517 Human sec
306	34	81.0	65	3	ABG03347	Abg03347 Human sec	379	34	81.0	520	6	ABR9517	Abr9517 Human sec
307	34	81.0	88	4	ABG03347	Abg03347 Novel hum	380	34	81.0	520	6	ABR9517	Abr9517 Human sec
308	34	81.0	94	4	ABG03347	Abg03347 Human col	381	34	81.0	520	6	ABR9517	Abr9517 Human sec
309	34	81.0	99	8	ABO59058	Abu59058 Human gen	382	34	81.0	520	6	ABR9517	Abr9517 Human sec
310	34	81.0	118	4	AAO04909	Aao04909 Human pol	383	34	81.0	520	6	ABR9517	Abr9517 Human sec
311	34	81.0	120	9	ADY28389	Ady28389 Novel Erb	384	34	81.0	520	6	ABR9517	Abr9517 Human sec
312	34	81.0	127	4	AAV74062	Aav74062 Human col	385	34	81.0	520	6	ABR9517	Abr9517 Human sec
313	34	81.0	143	7	ADK31516	Adk31516 Human nov	386	34	81.0	520	6	ABO00267	Abu000267 Novel hum
314	34	81.0	145	5	ADK34312	Adk34312 Novel hum	387	34	81.0	520	6	ABO00267	Abu000267 Novel hum
315	34	81.0	150	3	AAE58994	Aae58994 Breast an	388	34	81.0	520	6	ABO02244	Abu002244 Human sec
316	34	81.0	153	4	AAW93636	Aaw93636 Human pol	389	34	81.0	520	6	ABU88818	Abu88818 Novel hum

390	34	81.0	520	6	ABU83513	Human sec	463	34	81.0	520	6	ABU87651	Human PRO
391	34	81.0	520	6	ABO06314	Novel hum	464	34	81.0	520	6	ABU91519	Human PRO
392	34	81.0	520	6	ABR59350	Human sec	465	34	81.0	520	6	ABU84733	Human sec
393	34	81.0	520	6	ABO09412	Human sec	466	34	81.0	520	6	ABR69823	Human sec
394	34	81.0	520	6	ABO19276	Novel hum	467	34	81.0	520	6	ABU80200	Human PRO
395	34	81.0	520	6	ABO11294	Human sec	468	34	81.0	520	6	ABU93469	Human PRO
396	34	81.0	520	6	ABR66912	Human sec	469	34	81.0	520	6	ABO10022	Human sec
397	34	81.0	520	6	ABO16125	Human sec	470	34	81.0	520	6	ABO09107	Human sec
398	34	81.0	520	6	ABO13831	Human sec	471	34	81.0	520	6	ABU10675	Human sec
399	34	81.0	520	6	ABU09490	Human sec	472	34	81.0	520	6	ABU95684	Human PRO
400	34	81.0	520	6	ABU65734	Human sec	473	34	81.0	520	6	ABU96893	Novel hum
401	34	81.0	520	6	ABO07582	Human PRO	474	34	81.0	520	6	ABR70738	Human sec
402	34	81.0	520	6	ABO03769	Human sec	475	34	81.0	520	6	ABO05089	Novel hum
403	34	81.0	520	6	ABR67217	Human sec	476	34	81.0	520	6	ABO08497	Human sec
404	34	81.0	520	6	ABO15820	Human sec	477	34	81.0	520	6	ABO05704	Human sec
405	34	81.0	520	6	ABU56101	Human sec	478	34	81.0	520	6	ABR74093	Human sec
406	34	81.0	520	6	ABU61162	Human PRO	479	34	81.0	520	6	ABR95685	Human sec
407	34	81.0	520	6	ABU65429	Human PRO	480	34	81.0	520	6	ABR80982	Human sec
408	34	81.0	520	6	ABU95374	Novel hum	481	34	81.0	520	6	ABR81287	Human sec
409	34	81.0	520	6	ABU71277	Human PRO	482	34	81.0	520	6	ABM00983	Human sec
410	34	81.0	520	6	ABO07887	Human PRO	483	34	81.0	520	6	ABR88585	Human sec
411	34	81.0	520	6	ABR70128	Human sec	484	34	81.0	520	6	ABM77406	Human sec
412	34	81.0	520	6	ABR69461	Human sec	485	34	81.0	520	6	ABO28890	Human sec
413	34	81.0	520	6	ABO01602	Human PRO	486	34	81.0	520	6	ABO31635	Human sec
414	34	81.0	520	6	ABU81404	Human PRO	487	34	81.0	520	6	ABM08052	Human sec
415	34	81.0	520	6	ABR60201	Human sec	488	34	81.0	520	6	ABO40532	Human sec
416	34	81.0	520	6	ABR67936	Human sec	489	34	81.0	520	6	ABO35957	Human PRO
417	34	81.0	520	6	ABR65324	Human sec	490	34	81.0	520	6	ABO44096	Human PRO
418	34	81.0	520	6	ABR68546	Human sec	491	34	81.0	520	6	ADA78084	Human sec
419	34	81.0	520	6	ABR71958	Human sec	492	34	81.0	520	6	ABM24891	Human sec
420	34	81.0	520	6	ABU85438	Human PRO	493	34	81.0	520	6	ABO03159	Human sec
421	34	81.0	520	6	ABU89128	Human sec	494	34	81.0	520	6	ABR90415	Human sec
422	34	81.0	520	6	ABU83208	Human sec	495	34	81.0	520	6	ABM17329	Human sec
423	34	81.0	520	6	ABU95064	Novel hum	496	34	81.0	520	6	ABR95075	Human sec
424	34	81.0	520	6	ABU90612	Novel hum	497	34	81.0	520	6	ABR95380	Human sec
425	34	81.0	520	6	ABU84123	Human sec	498	34	81.0	520	6	ABO21618	Human sec
426	34	81.0	520	6	ABU93774	Novel hum	499	34	81.0	520	6	ABR97882	Human sec
427	34	81.0	520	6	ABR65019	Human sec	500	34	81.0	520	6	ABR87670	Human sec
428	34	81.0	520	6	ABR68851	Human sec							
429	34	81.0	520	6	ABO06667	Human sec							
430	34	81.0	520	6	ABR99212	Human sec							
431	34	81.0	520	6	ABU57096	Human PRO							
432	34	81.0	520	6	ABU86048	Novel hum							
433	34	81.0	520	6	ABU82335	Novel hum							
434	34	81.0	520	6	ABU87346	Human PRO							
435	34	81.0	520	6	ABU83818	Human sec							
436	34	81.0	520	6	ABO08192	Human PRO							
437	34	81.0	520	6	ABU81903	Novel hum							
438	34	81.0	520	6	ABU66067	Novel hum							
439	34	81.0	520	6	ABR59896	Human sec							
440	34	81.0	520	6	ABR94084	Novel hum							
441	34	81.0	520	6	ABU80431	Human sec							
442	34	81.0	520	6	ABU99937	Novel hum							
443	34	81.0	520	6	ABR66607	Human sec							
444	34	81.0	520	6	ABR91025	Human sec							
445	34	81.0	520	6	ABU94452	Human PRO							
446	34	81.0	520	6	ABU79334	Human PRO							
447	34	81.0	520	6	ABU86663	Human sec							
448	34	81.0	520	6	ABU86968	Novel hum							
449	34	81.0	520	6	ABU94757	Human PRO							
450	34	81.0	520	6	ABO04684	Human PRO							
451	34	81.0	520	6	ABR70433	Human sec							
452	34	81.0	520	6	ABU98598	Human PRO							
453	34	81.0	520	6	ABR65997	Human sec							
454	34	81.0	520	6	ABR64714	Human sec							
455	34	81.0	520	6	ABU79639	Human PRO							
456	34	81.0	520	6	ABU93030	Human sec							
457	34	81.0	520	6	ABU95989	Human PRO							
458	34	81.0	520	6	ABU91209	Novel hum							
459	34	81.0	520	6	ABU90302	Novel hum							
460	34	81.0	520	6	ABO09717	Human sec							
461	34	81.0	520	6	ABO10989	Human sec							
462	34	81.0	520	6	ABR71043	Human sec							

RESULT 1
ADOI5278
ID ADOI5278 standard; peptide; 8 AA.
XX
AC ADOI5278;
XX
DT 12-AUG-2004 (first entry)
XX
DE Collagen XIII and alpha1beta1 integrin binding inhibiting peptide SEQ:1.
XX
KW Chronic inflammatory disease; blocking agent;
KW Collagen XIII and alpha1beta1 integrin binding inhibitor; collagen XIII;
KW alpha1beta1 integrin; inflammatory disease; antiinflammatory;
KW antipsoriasis; antiarthritic; gene therapy; renal fibrosis;
KW lung fibrosis; liver fibrosis; rheumatoid arthritis; psoriasis;
XX experimental colitis; crescentic glomerulonephritis.
OS Synthetic.
XX
XX WO2004041846-A2.
PN
XX 21-MAY-2004.
PD
XX
XX 31-OCT-2003; 2003WO-US034818.
PF
XX
XX 01-NOV-2002; 2002US-0423297P.
PR
XX
XX (BOYS-) BOYS TOWN NAT RES HOSPITAL.
PA
XX
XX Cosgrove D;
PI


```
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Rioux P, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42172.
DR
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 652 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 652;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPCL 8
DB 291 GAEGSPGL 298
RESULT 4
ABM83519
ID ABM83519 standard; protein; 654 AA.
XX
XX AC ABM83519;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3768.
DE
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
OS
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410359P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI
```

```
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Rioux P, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42171.
DR
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 654 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 654;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
DB 293 GAEGSPGL 300
RESULT 5
ABM83518
ID ABM83518 standard; protein; 657 AA.
XX
XX AC ABM83518;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3767.
DE
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
OS
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
```



```
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kitton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42168.
DR
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 673 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 673;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
DB 312 GAEGSPGL 319
RESULT 8
ID ABM83515 standard; protein; 683 AA.
XX
XX ABM83515;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3764.
DE
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2004023973-A2.
PN
XX
XX 25-MAR-2004.
PD
XX
XX 12-SEP-2003; 2003WO-US028227.
PF
XX
XX 12-SEP-2002; 2002US-0410259P.
PR
XX
XX 12-SEP-2002; 2002US-0410260P.
PR
XX
XX (INCY-) INCYTE CORP.
PA
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eider LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kitton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42167.
DR
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 683 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 683;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
DB 334 GAEGSPGL 341
RESULT 9
ABM83514
ID ABM83514 standard; protein; 685 AA.
XX
XX ABM83514;
AC
XX
XX 18-NOV-2004 (first entry)
DT
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3763.
DE
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2004023973-A2.
PN
XX
XX 25-MAR-2004.
PD
XX
XX 12-SEP-2003; 2003WO-US028227.
PF
XX
XX 12-SEP-2002; 2002US-0410259P.
PR
XX
XX 12-SEP-2002; 2002US-0410260P.
PR
XX
XX (INCY-) INCYTE CORP.
PA
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eider LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
```

```
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42166.
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX Sequence 685 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 685;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 324 GAEGSPGL 331
RESULT 10
ID ABM83529 standard; protein; 688 AA.
XX ABM83529;
AC ABM83529;
XX 18-NOV-2004 (first entry)
DT Human diagnostic and therapeutic pprotein SEQ ID NO:3778.
DE gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX Homo sapiens.
OS WO2004023973-A2.
XX 25-MAR-2004.
PD 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
XX 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42181.
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX Sequence 688 AA;
SQ
Query Match 100.0%; Score 42; DB 8; Length 688;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 353 GAEGSPGL 360
RESULT 11
ID ABM83528 standard; protein; 692 AA.
XX ABM83528;
AC ABM83528;
XX 18-NOV-2004 (first entry)
DT Human diagnostic and therapeutic pprotein SEQ ID NO:3777.
DE gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX Homo sapiens.
OS WO2004023973-A2.
XX 25-MAR-2004.
PD 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
XX 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
```


PI Patury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
XX N-PSDB; ACN42180.
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX SQ Sequence 692 AA;
Query Match 100.0%; Score 42; DB 8; Length 692;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;
QY 1 GAEGSPGL 8
Db 331 GAEGSPGL 338
RESULT 12
ABM83527
ID ABM83527 standard; protein; 695 AA.
AC ABM83527;
XX 18-NOV-2004 (first entry)
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3776.
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX Homo sapiens.
XX WO2004023973-A2.
XX 25-MAR-2004.
XX 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
XX 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RU, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.
XX N-PSDB; ACN42179.
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX Claim 27; Page; 190pp; English.
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX SQ Sequence 695 AA;
Query Match 100.0%; Score 42; DB 8; Length 695;
Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;
QY 1 GAEGSPGL 8
Db 334 GAEGSPGL 341
RESULT 13
ABM83526
ID ABM83526 standard; protein; 697 AA.
AC ABM83526;
XX 18-NOV-2004 (first entry)
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3775.
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX Homo sapiens.
XX WO2004023973-A2.
XX 25-MAR-2004.
XX 12-SEP-2003; 2003WO-US028227.
XX 12-SEP-2002; 2002US-0410259P.
XX 12-SEP-2002; 2002US-0410260P.
XX (INCY-) INCYTE CORP.
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;

DR WPI; 2004-329368/30.
DR N-PSDB; ACN42178.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 697 AA;
Query Match 100.0%; Score 42; DB 8; Length 697;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 336 GAEGSPGL 343
RESULT 14
ABM83525
ID ABM83525 standard; protein; 699 AA.
XX
AC ABM83525;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3774.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-04102598.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Deleagane AM, Panesar SR, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
WPI; 2004-329368/30.
DR

DR N-PSDB; ACN42177.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 699 AA;
Query Match 100.0%; Score 42; DB 8; Length 699;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 353 GAEGSPGL 360
RESULT 15
ABM83524
ID ABM83524 standard; protein; 700 AA.
XX
AC ABM83524;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3773.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Deleagane AM, Panesar SR, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
WPI; 2004-329368/30.
DR

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.

XX Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm

XX Sequence 700 AA;

Query Match 100.0%; Score 42; DB 8; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
Db ||||| 353 GAEGSPGL 360

RESULT 16

ABM83522
ID ABM83522 standard; protein; 702 AA.

AC ABM83522;

DT 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:3771.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Fatury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.
DR N-PSDB; ACN42174.

XX

PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.

XX Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm

XX Sequence 702 AA;

Query Match 100.0%; Score 42; DB 8; Length 702;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
Db ||||| 353 GAEGSPGL 360

RESULT 17

ABM83523
ID ABM83523 standard; protein; 702 AA.

AC ABM83523;

DT 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:3772.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Fatury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42175.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful

DR WPI; 2004-329368/30.
 DR N-PSDB; ACN42173.
 XX
 PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.
 XX
 PS Claim 27; Page; 190pp; English.
 XX
 CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 714 AA;
 Query Match 100.0%; Score 42; DB 8; Length 714;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPGL 8
 Db 353 GAEGSPGL 360
 |||||
 RESULT 20
 ADN95515
 ID ADN95515 standard; protein; 717 AA.
 AC ADN95515;
 DT 01-JUL-2004 (first entry)
 DE Human BEC/LEC-related protein sequence SeqID438.
 XX
 XX growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytostatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.
 XX
 OS Homo sapiens.
 XX
 PN WO2003080640-A1.
 XX
 PD 02-OCT-2003.
 XX
 PF 07-MAR-2003; 2003WO-US006900.
 XX
 PR 07-MAR-2002; 2002US-0363019P.
 XX
 XX (LUDW-) LUDWIG INST CANCER RES.
 PA (LICN) LICENTIA LTD.
 XX
 XX Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 PI WPI; 2003-876899/81.
 XX DR N-PSDB; ADN95516.
 XX
 XX Example 1; SEQ ID NO 438; 176pp; English.

XX
 CC This invention relates to a method of differentially modulating the
 CC growth or differentiation of blood endothelial cells (BEC) or lymphatic
 CC endothelial cells (LEC) comprises contacting endothelial cells with a
 CC composition comprising an agent that differentially modulates blood or
 CC lymphatic endothelial cells. Treating hereditary lymphoedema comprises
 CC identifying a human subject with lymphoedema and with a mutation in at
 CC least one allele of a gene encoding a LEC protein, where the mutation
 CC correlates with lymphoedema in human subjects, and with the proviso that
 CC the LEC protein is not VEGFR-3; and administering to the subject a
 CC composition comprising a lymphatic growth agent selected from VEGF-C or
 CC VEGF-D polypeptides and polynucleotides. The invention may be useful for
 CC the development of compounds with an antiangiogenic, cytostatic,
 CC vasotropic or antiinflammatory activity or for gene therapy. The method
 CC is useful in modulating the growth or differentiation of blood
 CC endothelial cells or lymphatic endothelial cells, in treating hereditary
 CC lymphoedema, in screening for an endothelial cell disorder or
 CC predisposition to the disorder or in monitoring the efficacy or toxicity
 CC of a drug on endothelial cells. The agent is useful in manufacturing a
 CC medicament for the differential modulation of blood vessel endothelial
 CC cell or lymphatic vessel endothelial cell growth or differentiation. The
 CC lymphatic growth agent may also be used in manufacturing a medicament for
 CC the treatment of hereditary lymphoedema resulting from a mutation in a
 CC LEC gene or of other diseases involving the lymphatic vessels, such as
 CC various inflammatory diseases and cancer metastasis via the lymphatic
 CC system. The present sequence is that of a human LEC/BEC differentially
 CC expressed protein which is related to the method of the invention. Note:
 CC This sequence does not appear in the specification but was obtained by
 CC the indexer using the source data given in table 14 of the specification.
 XX
 SQ Sequence 717 AA;
 Query Match 100.0%; Score 42; DB 7; Length 717;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPGL 8
 Db 356 GAEGSPGL 363
 |||||
 RESULT 21
 ABO83203
 ID ABO83203 standard; protein; 495 AA.
 XX
 AC ABO83203;
 XX
 DT 29-JUL-2004 (first entry)
 XX
 DE Pseudomonas aeruginosa polypeptide #15378.
 XX
 KW Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN US6551795-B1.
 XX
 PD 22-APR-2003.
 XX
 PF 18-FEB-1999; 99US-00252991.
 XX
 PR 18-FEB-1998; 98US-0074788P.
 PR 27-JUL-1998; 98US-0094190P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA
 XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;
 PI WPI; 2003-615309/58.
 XX DR N-PSDB; ABD16774.
 XX
 XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of

PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 31949; 455pp; English.

XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-*P. aeruginosa* drugs, as templates for recombinant
 CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
 CC infection, and in detection of *P. aeruginosa* sequences or other sequences
 CC of *Pseudomonas* species using biochip technology. Sequences AB067826-
 CC AB084396 represent *P. aeruginosa* polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html

XX SQ Sequence 495 AA;

Query Match 92.9%; Score 39; DB 7; Length 495;
 Best Local Similarity 87.5%; Pred. No. 2.6e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 DB 208 GAEGSPGV 215

RESULT 22

ADW44464
 ID ADW44464 standard; protein; 738 AA.

XX AC ADW44464;

XX 24-MAR-2005 (first entry)

XX Murine collagen alpha-2(IV) chain DNA.

XX cell transduction; nerves; cell adhesion; collagen alpha-2(IV) chain.

XX Mus musculus.

XX WO2005001090-A1.

XX 06-JAN-2005.

XX 25-JUN-2004; 2004WO-JP009568.

XX 26-JUN-2003; 2003JP-00183630.

XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.

XX Miyake M, Uchimura E, Yoshikawa T, Miyake J;

XX WPI; 2005-091515/10.

XX N-PSDB; ADW44463.

XX Composition useful for improving transduction efficiency of nucleic acid
 PT into cell, comprises cell adhesion molecule and gene-transfer reagent.

XX Disclosure; SEQ ID NO 8; 446pp; Japanese.

XX This invention describes a novel composition for improving transduction
 CC efficiency of a nucleic acid into a cell from nervous tissue which
 CC comprises a cell adhesion molecule and a gene-transfer reagent. The
 CC invention also describes a device, kit and novel method for improving
 CC nucleic acid transduction efficiency of a cell on a solid phase. The cell
 CC adhesion molecule contains an extracellular matrix chosen from collagen,
 CC laminin and fibronectin. The gene transfer reagent comprises a cationic

CC polymer, cationic liquid, polyamine type reagent, polyamine type reagent
 CC or calcium phosphate. This sequence represents the murine collagen alpha-
 CC 2(IV) chain which is used in the composition of the invention.

XX SQ Sequence 738 AA;

Query Match 92.9%; Score 39; DB 9; Length 738;
 Best Local Similarity 87.5%; Pred. No. 3.8e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 DB 166 GAQGSPL 173

RESULT 23

AAV23942

ID AAV23942 standard; protein; 30 AA.

XX AC AAV23942;

XX 22-SEP-1999 (first entry)

XX Amino acid sequence of a collagen-like protein.

XX Collagen-like protein; promoter; *Bacillus brevis*.

XX Synthetic.

XX JPI1178574-A.

XX 06-JUL-1999.

XX 22-DEC-1997; 97JP-00353216.

XX 22-DEC-1997; 97JP-00353216.

XX (TOYW) TOYOTA CHUO KENKYUSHO KK.

XX (HGET) HIGETA SHOYU KK.

XX WPI; 1999-437307/37.

XX New collagen-like protein - useful for large scale production of stable
 PT collagen-like protein.

XX Example 2; Page 10; 13pp; Japanese.

XX The specification describes a recombinant DNA in which DNA encoding a
 CC collagen-like protein is linked to the 3' terminal of a DNA containing a
 CC promoter region derived from *Bacillus brevis*. The bacterium is cultured,
 CC and used for the useful for the large scale production of a stable
 CC soluble collagen-like protein in the medium. The present sequence
 CC represents a collagen-like protein of the invention

XX SQ Sequence 30 AA;

Query Match 90.5%; Score 38; DB 2; Length 30;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
 |||||

DB 4 GAEGSPG 10

RESULT 24

AAV23938

ID AAV23938 standard; protein; 30 AA.

XX AC AAV23938;

XX 22-SEP-1999 (first entry)

DE Amino acid sequence of a collagen-like protein.
XX Collagen-like protein; promoter; Bacillus brevis.
XX Synthetic.
XX JP11178574-A.
XX 06-JUL-1999.
XX 22-DEC-1997; 97JP-00353216.
XX 22-DEC-1997; 97JP-00353216.
XX (TOYU) TOYOTA CHUO KENKYUSHO KK.
XX (HGET) HIGETA SHOUYU KK.
XX WPI; 1999-437307/37.
XX New collagen-like protein - useful for large scale production of stable
PT collagen-like protein.
XX Claim 5; Page 7; 13pp; Japanese.
XX The specification describes a recombinant DNA in which DNA encoding a
CC collagen-like protein is linked to the 3' terminal of a DNA containing a
CC promoter region derived from Bacillus brevis. The bacterium is cultured,
CC and used for the useful for the large scale production of a stable
CC soluble collagen-like protein in the medium. The present sequence
CC represents a collagen-like protein of the invention
XX
SQ Sequence 30 AA;

Query Match 90.5%; Score 38; DB 2; Length 30;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGSPG 7
DB 10 GAGGSPG 16

RESULT 25
ABU53024
ID ABU53024 standard; protein; 134 AA.
XX AC ABU53024;
XX 14-APR-2003 (first entry)
XX Human testes-derived DKFZphtes3_18f3 homologue #20.
XX Homo sapiens.
XX WO200112659-A2.
XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB001496.
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX Wiemann S;
XX WPI; 2001-327840/34.
XX Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.

XX Example III; Page 637; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 134 AA;

Query Match 90.5%; Score 38; DB 4; Length 134;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGSPG 7
DB 97 GAGGSPG 103

RESULT 26
ABU53020
ID ABU53020 standard; protein; 136 AA.
XX AC ABU53020;
XX 14-APR-2003 (first entry)
XX Human testes-derived DKFZphtes3_18f3 homologue #16.
XX Homo sapiens.
XX WO200112659-A2.
XX 22-FEB-2001.
XX 18-AUG-2000; 2000WO-IB001496.
XX 18-AUG-1999; 99US-0149499P.
XX 28-SEP-1999; 99US-0156503P.
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.
XX Wiemann S;
XX WPI; 2001-327840/34.
XX Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.

XX Example III; Page 636; 1095pp; English.
XX This invention describes novel polynucleotides and polypeptides isolated
CC from human cDNA libraries which can be used for gene therapy or in
CC vaccines. The polynucleotides of the invention and antibodies encoded by
CC them may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The products of the
CC invention may also be used to identify modulators of expression and
CC activity and to down regulate expression and activity. The antibodies of
CC the invention may also be used as diagnostic agents for detecting the
CC presence of polypeptides in samples. This sequence represents a homologue
CC of a polypeptide described in the disclosure of the invention
XX
SQ Sequence 136 AA;

Query Match 90.5%; Score 38; DB 4; Length 136;

XX Nucleic acids having the sequences of clones isolated from libraries of
PT different human tissues, useful in recombinant DNA methodologies.

PA	(USUU) US SURGICAL CORP.
XX	
PI	Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX	
DR	WPI; 2000-259138/23.
XX	
PT	Production of extracellular matrix proteins containing 4-trans-
PT	hydroxyproline results in native self aggregating proteins, useful on
PT	medical implants.
XX	
PS	Claim 10; Fig 80; 260pp; English.
XX	
CC	The specification describes a method for producing an extracellular
CC	matrix protein or its fragment. The extracellular matrix protein is
CC	capable of self aggregating in a cell which does not ordinarily
CC	hydroxylated prolines. The method comprises optimising a nucleic acid
CC	sequence for expression in the cell by substitution of codons preferred
CC	by that cell for naturally occurring codons not preferred by the cell;
CC	incorporating the nucleic acid sequence into the cell; and contacting the
CC	cell with a hypertonic growth medium containing at least one amino acid,
CC	selected from the group consisting of trans-4-hydroxyproline and 3-
CC	hydroxyproline to allow at least one of the amino acids to be assimilated
CC	into the cell and incorporated into the extracellular matrix protein. The
CC	method may be used to make host cells assimilate and incorporate trans-4-
CC	hydroxyproline into proteins. This is especially useful in the
CC	recombinant production of proteins such as collagen, fibrinogen and
CC	fibronectin whose ability to self aggregate and produce functional
CC	proteins depends on the post translational hydroxylation of proline. The
CC	method is also useful in studying the structure and function of
CC	polypeptides which do not normally contain trans-4-hydroxyproline. The
CC	present sequence represents a C-terminal fragment of human collagen type
CC	1 (alpha2), with optimised codon usage, designated D4
XX	
SQ	Sequence 219 AA;
	Query Match 90.5%; Score 38; DB 3; Length 219;
	Best Local Similarity 100.0%; Pred. No. 1.7e+02;
	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GAEGSPG 7
Db	20 GAEGSPG 26
	RESULT 30
ID	AAY84553
ID	AAY84553 standard; protein; 219 AA.
XX	
AC	AAY84553;
XX	
DT	25-JUL-2000 (first entry)
DE	A C-terminal fragment of human collagen type 1 (alpha).
XX	
KW	Extracellular matrix protein; self aggregation; hydroxylated proline;
KW	trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
KW	collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX	
OS	Homo sapiens.
XX	
PN	EP992586-A2.
XX	
PD	12-APR-2000.
XX	
PF	07-OCT-1999; 99EP-00119184.
XX	
PR	09-OCT-1998; 98US-00169768.
XX	
PA	(USUU) US SURGICAL CORP.
XX	
PI	Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX	
DR	WPI; 2000-259138/23.
XX	
PT	Production of extracellular matrix proteins containing 4-trans-
PT	hydroxyproline results in native self aggregating proteins, useful on
PT	medical implants.
XX	
N-PSDB; AAA12518.	
XX	
PT	Production of extracellular matrix proteins containing 4-trans-
PT	hydroxyproline results in native self aggregating proteins, useful on
PT	medical implants.
XX	
PS	Claim 16; Fig 72; 260pp; English.
XX	
CC	The specification describes a method for producing an extracellular
CC	matrix protein or its fragment. The extracellular matrix protein is
CC	capable of self aggregating in a cell which does not ordinarily
CC	hydroxylated prolines. The method comprises optimising a nucleic acid
CC	sequence for expression in the cell by substitution of codons preferred
CC	by that cell for naturally occurring codons not preferred by the cell;
CC	incorporating the nucleic acid sequence into the cell; and contacting the
CC	cell with a hypertonic growth medium containing at least one amino acid,
CC	selected from the group consisting of trans-4-hydroxyproline and 3-
CC	hydroxyproline to allow at least one of the amino acids to be assimilated
CC	into the cell and incorporated into the extracellular matrix protein. The
CC	method may be used to make host cells assimilate and incorporate trans-4-
CC	hydroxyproline into proteins. This is especially useful in the
CC	recombinant production of proteins such as collagen, fibrinogen and
CC	fibronectin whose ability to self aggregate and produce functional
CC	proteins depends on the post translational hydroxylation of proline. The
CC	method is also useful in studying the structure and function of
CC	polypeptides which do not normally contain trans-4-hydroxyproline. The
CC	present sequence represents a C-terminal fragment of human collagen type
CC	1 (alpha2), with optimised codon usage, designated D4
XX	
SQ	Sequence 219 AA;
	Query Match 90.5%; Score 38; DB 3; Length 219;
	Best Local Similarity 100.0%; Pred. No. 1.7e+02;
	Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GAEGSPG 7
Db	20 GAEGSPG 26
	RESULT 31
ID	ABO67957
ID	ABO67957 standard; protein; 238 AA.
XX	
AC	ABO67957;
XX	
DT	29-JUL-2004 (first entry)
XX	
XX	Pseudomonas aeruginosa polypeptide #132.
XX	Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
OS	Pseudomonas aeruginosa.
XX	
PN	US6551795-B1.
XX	
PD	22-APR-2003.
XX	
PF	18-FEB-1999; 99US-00252991.
XX	
PR	18-FEB-1998; 98US-0074788P.
PR	27-JUL-1998; 98US-0094190P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX	
DR	WPI; 2003-615309/58.
DR	N-PSDB; ABD01528.
XX	
PT	Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide, f
PT	useful as molecular targets for diagnostics, prophylaxis and treatment of
PT	pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 16703; 455pp; English.

PS The invention relates to *Pseudomonas aeruginosa* polypeptides and the

XX polynucleotides encoding them. The sequences are useful in diagnosis and

CC therapy of pathological conditions, as molecular targets for diagnostics,

CC prophylaxis and treatment of pathological conditions resulting from a

CC bacterial infection, for evaluating a compound, such as a polypeptide,

CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of

CC effective antibacterial targets, as targets for antibacterial drugs,

CC including anti-*P. aeruginosa* drugs, as templates for recombinant

CC production of *P. aeruginosa*-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of *P. aeruginosa*-caused

CC infection, and in detection of *P. aeruginosa* sequences or other sequences

CC of *Pseudomonas* species using biochip technology. Sequences AB067826-

CC AB084396 represent *P. aeruginosa* polypeptides of the invention. Note: The

CC sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format from USPTO at

CC seqdata.uspto.gov/sequence.html

XX SQ Sequence 238 AA;

Query Match 90.5%; Score 38; DB 7; Length 238;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEKSPG 7

Db 56 GAEKSPG 62

|||||

RESULT 32

ABE27174

ID ABE27174 standard; protein; 317 AA.

XX AC ABE27174;

XX 22-SEP-2005 (first entry)

DE Pinus radiata cell cycle protein SEQ ID NO 335.

XX plant protectant; fungicide; plant growth regulant; gene therapy;

KW cell cycle; gene expression; plant; transgenic plant; microarray; wood.

XX Pinus radiata.

OS WO2005065339-A2.

PN 21-JUL-2005.

XX 30-DEC-2004; 2004WO-US043804.

PF 30-DEC-2003; 2003US-0533036P.

XX (ARBO-) ARBOGEN LLC.

PA (GENE-) GENESIS RES & DEV CORP LTD.

XX Forster RL, Connett MB, Emerson SJ, Grigor MR, Higgins CM;

PI Lund ST, Magusin A, Kodrzycki RJ;

XX WPI; 2005-506765/51.

DR N-PSDB; ABE26914.

XX New polynucleotide encoding a plant cell cycle protein, useful for

PT modifying plant development and altering plant phenotype.

XX Claim 24; SEQ ID NO 335; 499pp; English.

XX The invention describes an isolated polynucleotide comprising: (i) a

CC sequence of SEQ ID NOS: 1-237 or their conservative variants; (ii) a

CC sequence encoding the catalytic or substrate-binding domain of a

CC polypeptide of SEQ ID NOS: 261-497, where the polynucleotide encodes a

CC polypeptide having the activity of the polypeptide of SEQ ID NOS: 261-497

CC ; or (iii) a nucleic acid sequence of SEQ ID NOS: 471-697. Also described

CC are: (1) a DNA construct comprising at least one polynucleotide having

CC the sequence of SEQ ID NOS: 1-237 or their conservative variants; (2) a

CC plant cell transformed with the DNA construct of (1); (3) a transgenic

CC plant comprising the plant cell of (2); (4) a method of making a

CC transformed plant; (5) a wood or a wood pulp obtained from a transgenic

CC tree which has been transformed with the DNA construct of (1); (6) a

CC method of making wood or wood pulp; (7) an isolated polypeptide

CC comprising an amino acid sequence encoded by the new isolated

CC polynucleotide or comprising any of the amino acid sequences of SEQ ID

CC NOS: 261-497; (8) a method of altering a plant phenotype of a plant; (9)

CC a method of correlating gene expression in two different samples; (10) a

CC method of correlating the possession of a plant phenotype to the level of

CC gene expression in the plant of one or more genes; (11) a method of

CC correlating gene expression to a stage of the cell cycle; (12) a

CC combination for detecting expression of one or more genes, comprising

CC two or more oligonucleotides, where each oligonucleotide is capable of

CC hybridizing to a nucleic acid sequence of SEQ ID NOS: 1-237 or to gene

CC product encoded by a nucleic acid sequence of SEQ ID NOS: 1-237; (13) a

CC microarray comprising the combination of (12) provided on a solid

CC support, where each of the two or more oligonucleotides occupies a unique

CC location on the solid support; (14) a method for detecting one or more

CC genes in a sample; (15) a method for detecting one or more nucleic acid

CC sequences encoded by one or more genes in a sample; and (16) a kit, for

CC detecting gene expression, comprising the microarray of (13) together

CC with one or more buffers or reagents for a nucleotide hybridization

CC reaction. The polynucleotides, polypeptides, DNA construct, composition,

CC and methods are useful for modifying plant development and altering plant

CC phenotype. This is the amino acid sequence of a cell cycle protein

CC isolated in the invention.

XX SQ Sequence 317 AA;

Query Match 90.5%; Score 38; DB 9; Length 317;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEKSPG 7

Db 311 GAEKSPG 317

|||||

RESULT 33

ADE87049

ID ADE87049 standard; protein; 465 AA.

XX AC ADE87049;

XX 29-JAN-2004 (first entry)

DE Human pancreatic cell protein sequence SeqID509.

XX neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;

KW cancer death; cytostatic; vaccine; gene therapy;

KW non-cancerous pancreas disease; human.

XX Homo sapiens.

OS WO2003060145-A2.

PN 24-JUL-2003.

XX 19-DEC-2002; 2002WO-US040655.

PF 21-DEC-2001; 2001US-0342768P.

XX (DIAD-) DIADEXUS INC.

XX Sun Y, Liu C;

PI WPI; 2003-587286/55.

XX N-PSDB; ADE87386.

PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
XX non-cancerous disease states of the pancreas.

PS Claim 12; SEQ ID NO 509; 635pp; English.

XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.

SQ Sequence 465 AA;

Query Match 90.5%; Score 38; DB 7; Length 465;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||||
DB 20 GAEGSPG 26

RESULT 34
ADW99569
ID ADW99569 standard; protein; 498 AA.

XX AC ADW99569;

XX DT 21-APR-2005 (first entry)

XX DE Human gelatin protein - SEQ ID 2.

XX KW gelatin; biofilm.

XX OS Homo sapiens.

XX PN WO2005012356-A2.

XX PD 10-FEB-2005.

XX PF 30-JUL-2004; 2004WO-US024663.

XX PR 01-AUG-2003; 2003US-0492085P.

XX PR 29-JUL-2004; 2004US-00901816.

XX PA (FIBR-) FIBROGEN INC.

XX PI Chang RC, Olsen DR, Polarek JW, Williams KE;

XX DR WPI; 2005-132654/14.

XX PT Recombinant gelatin for capsule manufacture.

XX PS Claim 25; SEQ ID NO 2; 74pp; English.

XX The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.

SQ Sequence 498 AA;

Query Match 90.5%; Score 38; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7

DB 487 GAEGSPG 493
|||||||

RESULT 35

ADW99568
ID ADW99568 standard; protein; 498 AA.

XX AC ADW99568;

XX DT 21-APR-2005 (first entry)

XX DE Human gelatin protein - SEQ ID 1.

XX KW gelatin; biofilm.

XX OS Homo sapiens.

XX PN WO2005012356-A2.

XX PD 10-FEB-2005.

XX PF 30-JUL-2004; 2004WO-US024663.

XX PR 01-AUG-2003; 2003US-0492085P.

XX PR 29-JUL-2004; 2004US-00901816.

XX PA (FIBR-) FIBROGEN INC.

XX PI Chang RC, Olsen DR, Polarek JW, Williams KE;

XX DR WPI; 2005-132654/14.

XX PT Recombinant gelatin for capsule manufacture.

XX PS Claim 25; SEQ ID NO 1; 74pp; English.

XX The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.

SQ Sequence 498 AA;

Query Match 90.5%; Score 38; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7

DB 487 GAEGSPG 493
|||||||

RESULT 36

ADE87046
ID ADE87046 standard; protein; 516 AA.

XX AC ADE87046;

XX DT 29-JAN-2004 (first entry)

XX DE Human pancreatic cell protein sequence SeqID506.

XX KW neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
XX cancer death; cytostatic; vaccine; gene therapy;
XX non-cancerous pancreas disease; human.

XX OS Homo sapiens.

XX PN WO2003060145-A2.

XX PD 24-JUL-2003.

XX QY

PP 19-DEC-2002; 2002WO-US040655.
XX
PR 21-DEC-2001; 2001US-0342768P.
XX
PA (DIAD-) DIADEXUS INC.
XX
PI Sun Y, Liu C;
XX
XX WPI; 2003-587286/55.
DR N-PSDB; ADE87380.
XX
XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
PT non-cancerous disease states of the pancreas.
XX
XX Claim 12; SEQ ID NO 506; 635pp; English.
XX
XX This invention relates to novel nucleic acids and proteins present in
CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
CC cause of cancer death worldwide, therefore accurate methods of diagnosis
CC and treatment are required. Compounds which modulate the proteins of the
CC invention may have cytostatic activity and the protein and DNA sequences
CC of the invention may be useful for the development of a vaccine or in
CC gene therapy. The composition and methods are useful in diagnosing,
CC staging, imaging, monitoring, preventing or treating pancreatic cancer
CC and non-cancerous disease states of the pancreas. The present sequence is
CC that of a human pancreatic protein of the invention.
XX
SQ Sequence 516 AA;

Query Match 90.5%; Score 38; DB 7; Length 516;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 291 GAEGSPG 297
|||||

RESULT 37
ADW99570
ID ADW99570 standard; protein; 660 AA.
XX
AC ADW99570;
XX
XX 21-APR-2005 (first entry)
XX
XX Human gelatin protein - SEQ ID 3.
XX
XX gelatin; biofilm.
XX
XX Homo sapiens.
XX
XX WO2005012356-A2.
XX
XX 10-FEB-2005.
XX
XX 30-JUL-2004; 2004WO-US024663.
XX
XX 01-AUG-2003; 2003US-0492085P.
XX
XX 29-JUL-2004; 2004US-00901816.
XX
XX (FIBR-) FIBROGEN INC.
XX
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
XX WPI; 2005-132654/14.
XX
XX Recombinant gelatin for capsule manufacture.
XX
XX Claim 25; SEQ ID NO 3; 74pp; English.
XX
XX The invention comprises the amino acid sequences of 13 human gelatin
XX proteins which are useful in the manufacture of gelatin capsules,
XX encapsulants, or films. The present amino acid sequence represents a
XX human gelatin protein of the invention.
XX
SQ Sequence 660 AA;

Query Match 90.5%; Score 38; DB 9; Length 660;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493
|||||

RESULT 38
ADW99571
ID ADW99571 standard; protein; 660 AA.
XX
AC ADW99571;
XX
XX 21-APR-2005 (first entry)
XX
XX Human gelatin protein - SEQ ID 4.
XX
XX gelatin; biofilm.
XX
XX Homo sapiens.
XX
XX WO2005012356-A2.
XX
XX 10-FEB-2005.
XX
XX 30-JUL-2004; 2004WO-US024663.
XX
XX 01-AUG-2003; 2003US-0492085P.
XX
XX 29-JUL-2004; 2004US-00901816.
XX
XX (FIBR-) FIBROGEN INC.
XX
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
XX WPI; 2005-132654/14.
XX
XX Recombinant gelatin for capsule manufacture.
XX
XX Claim 25; SEQ ID NO 4; 74pp; English.
XX
XX The invention comprises the amino acid sequences of 13 human gelatin
XX proteins which are useful in the manufacture of gelatin capsules,
XX encapsulants, or films. The present amino acid sequence represents a
XX human gelatin protein of the invention.
XX
SQ Sequence 660 AA;

Query Match 90.5%; Score 38; DB 9; Length 660;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493
|||||

RESULT 39
ADR08580
ID ADR08580 standard; protein; 773 AA.
XX
AC ADR08580;
XX
XX 04-NOV-2004 (first entry)
XX
XX Human protein useful for treating neurological disease Seq 2086.
XX

XX human; oligo-capping method; diagnostic marker; gene therapy;
 KW osteoporosis; neurological disease; Alzheimer's disease;
 KW Parkinson's disease; dementia; short memory; cancer;
 KW sense or motor function; emotional reaction; fear response; panic;
 KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
 KW tranquiliser.
 XX Homo sapiens.
 XX
 XX
 PN EP1447413-A2.
 XX
 PD 18-AUG-2004.
 XX
 XX 12-FEB-2004; 2004EP-00003145.
 PF
 XX 14-FEB-2003; 2003JP-00102207.
 PR 09-MAY-2003; 2003JP-00131452.
 XX
 XX (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;
 XX
 DR WPI; 2004-583265/57.
 DR N-PSDB; ADR06624.
 XX
 XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
 XX
 PS Claim 1; SEQ ID NO 2086; 2686pp; English.
 XX
 CC This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention
 CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cytostatic and tranquiliser activities. This polypeptide is a protein
 CC encoded by a full length human cDNA sequence of the invention. NOTE: This
 CC sequence is not given in the sequence listing of the specification but
 CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
 CC office.
 XX
 SQ Sequence 773 AA;
 Query Match 90.5%; Score 38; DB 8; Length 773;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPG 7
 Db 260 GAEGSPG 266
 RESULT 40
 ADE87052
 ID ADE87052 standard; protein; 886 AA.
 XX
 AC ADE87052;
 XX
 XX 29-JAN-2004 (first entry)
 DT
 XX Human pancreatic cell protein sequence seqID512.
 DE
 XX

KW neoplastic pancreatic cell; pancreatic cancer; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 KW non-cancerous pancreas disease; human.
 XX Homo sapiens.
 OS
 PN WO2003060145-A2.
 XX
 PD 24-JUL-2003.
 XX
 XX 19-DEC-2002; 2002WO-US040655.
 PF
 XX 21-DEC-2001; 2001US-0342768P.
 PR
 XX (DIAD-) DIADEXUS INC.
 PA
 XX Sun Y, Liu C;
 PI
 XX WPI; 2003-587286/55.
 DR N-PSDB; ADE87389.
 DR
 XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.
 XX
 XX Claim 12; SEQ ID NO 512; 635pp; English.
 PS
 CC This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.
 XX
 SQ Sequence 886 AA;
 Query Match 90.5%; Score 38; DB 7; Length 886;
 Best Local Similarity 100.0%; Pred. No. 6.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPG 7
 Db 441 GAEGSPG 447
 RESULT 41
 ABO83805
 ID ABO83805 standard; protein; 926 AA.
 XX
 AC ABO83805;
 XX
 XX 29-JUL-2004 (first entry)
 DT
 XX Pseudomonas aeruginosa polypeptide #15980.
 DE
 XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.
 KW
 KW Pseudomonas aeruginosa.
 OS
 XX US6551795-B1.
 PN
 XX 22-APR-2003.
 PD
 XX 18-FEB-1999; 99US-00252991.
 PF
 XX 18-FEB-1998; 98US-0074788P.
 PR 27-JUL-1998; 98US-0094190P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA


```
DE Human gelatin protein - SEQ ID 7.
KW gelatin; biofilm.
XX
OS Homo sapiens.
XX
PN WO2005012356-A2.
XX
PD 10-FEB-2005.
XX
PF 30-JUL-2004; 2004WO-US024663.
XX
PR 01-AUG-2003; 2003US-0492085P.
XX
PR 29-JUL-2004; 2004US-00901816.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
DR WPI; 2005-132654/14.
XX
PT Recombinant gelatin for capsule manufacture.
XX
PS Claim 25; SEQ ID NO 7; 74pp; English.
XX
CC The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.
XX
SQ Sequence 1014 AA;

Query Match          90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
DB 841 GAEGSPG 847

RESULT 46
ADW99573
ID ADW99573 standard; protein; 1014 AA.
XX
AC ADW99573;
XX
DT 21-APR-2005 (first entry)
XX
DE Human gelatin protein - SEQ ID 6.
XX
KW gelatin; biofilm.
XX
OS Homo sapiens.
XX
PN WO2005012356-A2.
XX
PD 10-FEB-2005.
XX
PF 30-JUL-2004; 2004WO-US024663.
XX
PR 01-AUG-2003; 2003US-0492085P.
XX
PR 29-JUL-2004; 2004US-00901816.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
DR WPI; 2005-132654/14.
XX
PT Recombinant gelatin for capsule manufacture.
XX
PS Claim 25; SEQ ID NO 6; 74pp; English.
XX
CC The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.
XX
SQ Sequence 1014 AA;

Query Match          90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
DB 841 GAEGSPG 847

RESULT 47
ADW99572
ID ADW99572 standard; protein; 1014 AA.
XX
AC ADW99572;
XX
DT 21-APR-2005 (first entry)
```

```
DE Human gelatin protein - SEQ ID 7.
KW gelatin; biofilm.
XX
OS Homo sapiens.
XX
PN WO2005012356-A2.
XX
PD 10-FEB-2005.
XX
PF 30-JUL-2004; 2004WO-US024663.
XX
PR 01-AUG-2003; 2003US-0492085P.
XX
PR 29-JUL-2004; 2004US-00901816.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
DR WPI; 2005-132654/14.
XX
PT Recombinant gelatin for capsule manufacture.
XX
PS Claim 25; SEQ ID NO 7; 74pp; English.
XX
CC The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.
XX
SQ Sequence 1014 AA;

Query Match          90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
DB 841 GAEGSPG 847

RESULT 45
ADW99577
ID ADW99577 standard; protein; 1014 AA.
XX
AC ADW99577;
XX
DT 21-APR-2005 (first entry)
XX
DE Human gelatin protein - SEQ ID 10.
XX
KW gelatin; biofilm.
XX
OS Homo sapiens.
XX
PN WO2005012356-A2.
XX
PD 10-FEB-2005.
XX
PF 30-JUL-2004; 2004WO-US024663.
XX
PR 01-AUG-2003; 2003US-0492085P.
XX
PR 29-JUL-2004; 2004US-00901816.
XX
PA (FIBR-) FIBROGEN INC.
XX
PI Chang RC, Olsen DR, Polarek JW, Williams KE;
XX
DR WPI; 2005-132654/14.
XX
PT Recombinant gelatin for capsule manufacture.
XX
PS Claim 25; SEQ ID NO 10; 74pp; English.
```

```

XX Human gelatin protein - SEQ ID 5.
DE gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 29-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.
XX Claim 25; SEQ ID NO 5; 74pp; English.
XX The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.
XX Sequence 1014 AA;
XX Query Match 90.5%; Score 38; DB 9; Length 1014;
XX Best Local Similarity 100.0%; Pred. No. 7.8e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 841 GAEGSPG 847

RESULT 48
ADW99576
ID ADW99576 standard; protein; 1014 AA.
XX AC ADW99576;
XX 21-APR-2005 (first entry)
XX Human gelatin protein - SEQ ID 9.
DE gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 29-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.

Query Match 90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 841 GAEGSPG 847

RESULT 49
ADW99578
ID ADW99578 standard; protein; 1014 AA.
XX AC ADW99578;
XX 21-APR-2005 (first entry)
XX Human gelatin protein - SEQ ID 11.
DE gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 29-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.

Claim 25; SEQ ID NO 11; 74pp; English.
The invention comprises the amino acid sequences of 13 human gelatin
proteins which are useful in the manufacture of gelatin capsules, a
encapsulants, or films. The present amino acid sequence represents a
human gelatin protein of the invention.
Sequence 1014 AA;
Query Match 90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 841 GAEGSPG 847

RESULT 50
ADW99580
ID ADW99580 standard; protein; 1014 AA.
XX AC ADW99580;
XX Recombinant gelatin for capsule manufacture.

```

```

PS Claim 25; SEQ ID NO 9; 74pp; English.
XX The invention comprises the amino acid sequences of 13 human gelatin
CC proteins which are useful in the manufacture of gelatin capsules,
CC encapsulants, or films. The present amino acid sequence represents a
CC human gelatin protein of the invention.
XX Sequence 1014 AA;
XX Query Match 90.5%; Score 38; DB 9; Length 1014;
XX Best Local Similarity 100.0%; Pred. No. 7.8e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 841 GAEGSPG 847

RESULT 49
ADW99578
ID ADW99578 standard; protein; 1014 AA.
XX AC ADW99578;
XX 21-APR-2005 (first entry)
XX Human gelatin protein - SEQ ID 11.
DE gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 29-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.

Claim 25; SEQ ID NO 11; 74pp; English.
The invention comprises the amino acid sequences of 13 human gelatin
proteins which are useful in the manufacture of gelatin capsules, a
encapsulants, or films. The present amino acid sequence represents a
human gelatin protein of the invention.
Sequence 1014 AA;
Query Match 90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 841 GAEGSPG 847

RESULT 50
ADW99580
ID ADW99580 standard; protein; 1014 AA.
XX AC ADW99580;
XX Recombinant gelatin for capsule manufacture.

```

```

DT 21-APR-2005 (first entry)
XX Human gelatin protein - SEQ ID 13.
XX gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 23-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.
XX Claim 25; SEQ ID NO 13; 74pp; English.
XX The invention comprises the amino acid sequences of 13 human gelatin
XX proteins which are useful in the manufacture of gelatin capsules,
XX encapsulants, or films. The present amino acid sequence represents a
XX human gelatin protein of the invention.
XX Sequence 1014 AA;
Query Match 90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
DB 841 GAEGSPG 847
RESULT 51
AA184544
ID AAY84544 standard; protein; 1057 AA.
XX AAY84544;
XX 25-JUL-2000 (first entry)
XX A human collagen 1 (alpha1) protein helical region.
XX Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX Homo sapiens.
XX EP992586-A2.
XX 12-APR-2000.
XX 07-OCT-1999; 99EP-00119184.
XX 09-OCT-1998; 98US-00169768.
XX (USSU ) US SURGICAL CORP.
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAA12503.
XX
XX 21-APR-2005 (first entry)
XX Human gelatin protein - SEQ ID 13.
XX gelatin; biofilm.
XX Homo sapiens.
XX WO2005012356-A2.
XX 10-FEB-2005.
XX 30-JUL-2004; 2004WO-US024663.
XX 01-AUG-2003; 2003US-0492085P.
XX 23-JUL-2004; 2004US-00901816.
XX (FIBR-) FIBROGEN INC.
XX Chang RC, Olsen DR, Polarek JW, Williams KE;
XX WPI; 2005-132654/14.
XX Recombinant gelatin for capsule manufacture.
XX Claim 25; SEQ ID NO 13; 74pp; English.
XX The invention comprises the amino acid sequences of 13 human gelatin
XX proteins which are useful in the manufacture of gelatin capsules,
XX encapsulants, or films. The present amino acid sequence represents a
XX human gelatin protein of the invention.
XX Sequence 1014 AA;
Query Match 90.5%; Score 38; DB 9; Length 1014;
Best Local Similarity 100.0%; Pred. No. 7.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
DB 841 GAEGSPG 847
RESULT 51
AA184544
ID AAY84544 standard; protein; 1057 AA.
XX AAY84544;
XX 25-JUL-2000 (first entry)
XX A human collagen 1 (alpha1) protein helical region.
XX Extracellular matrix protein; self aggregation; hydroxylated proline;
XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
XX collagen; fibrinogen; fibronectin; post translational hydroxylation.
XX Homo sapiens.
XX EP992586-A2.
XX 12-APR-2000.
XX 07-OCT-1999; 99EP-00119184.
XX 09-OCT-1998; 98US-00169768.
XX (USSU ) US SURGICAL CORP.
XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
XX WPI; 2000-259138/23.
XX N-PSDB; AAA12503.
XX
XX Production of extracellular matrix proteins containing 4-trans-
XX hydroxyproline results in native self aggregating proteins, useful on
XX medical implants.
XX Example 10; Fig 39A-E; 260pp; English.
XX The specification describes a method for producing an extracellular
XX matrix protein or its fragment. The extracellular matrix protein is
XX capable of self aggregating in a cell which does not ordinarily
XX hydroxylated prolines. The method comprises optimising a nucleic acid
XX sequence for expression in the cell by substitution of codons preferred
XX by that cell for naturally occurring codons not preferred by the cell;
XX incorporating the nucleic acid sequence into the cell; and contacting the
XX cell with a hypertonic growth medium containing at least one amino acid,
XX selected from the group consisting of trans-4-hydroxyproline and 3-
XX hydroxyproline to allow at least one of the amino acids to be assimilated
XX into the cell and incorporated into the extracellular matrix protein. The
XX method may be used to make host cells assimilate and incorporate trans-4-
XX hydroxyproline into proteins. This is especially useful in the
XX recombinant production of proteins such as collagen, fibrinogen and
XX fibronectin whose ability to self aggregate and produce functional
XX proteins depends on the post translational hydroxylation of proline. The
XX method is also useful in studying the structure and function of
XX polypeptides which do not normally contain trans-4-hydroxyproline. The
XX present sequence represents human collagen 1 (alpha1) helical region,
XX which may be produced using the method of the invention
XX Sequence 1057 AA;
Query Match 90.5%; Score 38; DB 3; Length 1057;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
DB 858 GAEGSPG 864
RESULT 52
AA184403
ID AAY84403 standard; protein; 1058 AA.
XX AAY84403;
XX 12-JUL-2000 (first entry)
XX Amino acid sequence of human type 1 (alpha1) collagen polypeptide.
XX Alpha1 collagen; 3,4-dehydro-L-proline; epoxidation; 3,4-epoxyproline;
XX collagen; muscel adhesive protein; bioadhesive.
XX Homo sapiens.
XX WO200014201-A1.
XX 16-MAR-2000.
XX 07-SEP-1999; 99WO-US020462.
XX 09-SEP-1998; 98US-0099652P.
XX (USSU ) US SURGICAL CORP.
XX (PAOL/) PAOLELLA D N.
XX (GRUS/) GRUSKIN E A.
XX (BUEC/) BUECHTER D D.
XX Paolella DN, Gruskin EA, Buechter DD;
XX WPI; 2000-271051/23.
XX N-PSDB; AAZ99843.
XX Incorporating non-natural amino acid into polypeptide, useful e.g. for

```

PT production of bioadhesives, by epoxidation or substitution of
 XX dehydroproline residues.

PS Disclosure; Fig 6; 66pp; English.

XX The present sequence represents a human type 1 (alpha) collagen protein.
 CC Peptides derived from the protein were used to demonstrate incorporation
 CC of 3,4-dehydro-L-proline into the peptide, using the method of the
 CC invention. The specification describes a method for the incorporation of
 CC non-natural amino acid into a polypeptide. The method comprises reacting
 CC at least one 3,4-dehydroproline residue in the polypeptide with an
 CC epoxidation reagent from a polypeptide containing at least one 3,4-
 CC epoxypoline residue. The method is used for studying the effects of non-
 CC natural amino acids on structure and function of polypeptides. The method
 CC is also useful for commercial production of collagen or mussel adhesive
 CC proteins (which are useful as bioadhesives), and for incorporating a wide
 CC variety of groups, including therapeutic ligands and biological probes,
 CC into polypeptides

XX Sequence 1058 AA;

Query Match 90.5%; Score 38; DB 3; Length 1058;
 Best Local Similarity 100.0%; Pred. No. 8.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 859 GAEGSPG 865

RESULT 53

AA84540
 ID AA84540 standard; protein; 1107 AA.

AC AA84540;

XX 25-JUL-2000 (first entry)

XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.

XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibronectin; fibronectin; post translational hydroxylation;
 KW decorin; chimera.

XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.

XX Key Location/Qualifiers

FT Misc-difference 858 /note= "Gly encoded by GCT"

XX EP992586-A2.

XX 12-APR-2000.

XX 07-OCT-1999; 99EP-00119184.

XX 09-OCT-1998; 98US-00169768.

XX (USSU) US SURGICAL CORP.

XX Grushkin EA, Buechter DD, Zhang G, Connolly K;

XX WPI; 2000-259138/23.

XX N-PSDB; AA12500.

XX Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.

XX Claim 24; Fig 18; 260pp; English.

XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention

XX Sequence 1107 AA;

Query Match 90.5%; Score 38; DB 3; Length 1107;
 Best Local Similarity 100.0%; Pred. No. 8.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 858 GAEGSPG 864

RESULT 54

ADE87050
 ID ADE87050 standard; protein; 1161 AA.

XX ADE87050;

XX 29-JAN-2004 (first entry)

XX Human pancreatic cell protein sequence SeqID510.

XX neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;

XX cancer death; cytostatic; vaccine; gene therapy;

XX non-cancerous pancreas disease; human.

XX Homo sapiens.

XX WO2003060145-A2.

XX 24-JUL-2003.

XX 19-DEC-2002; 2002WO-US040655.

XX 21-DEC-2001; 2001US-0342768P.

XX (DIAD-) DIADEXUS INC.

XX Sun Y, Liu C;

XX WPI; 2003-587286/55.

XX N-PSDB; ADE87387.

XX New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.

XX Claim 12; SEQ ID NO 510; 635pp; English.

XX This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 CC cause of cancer death worldwide, therefore accurate methods of diagnosis

CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.

XX Sequence 1161 AA;

Query Match 90.5%; Score 38; DB 7; Length 1161;
 Best Local Similarity 100.0%; Pred. No. 8.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEKSPG 7
 |||||
 Db 716 GAEKSPG 722

RESULT 55
 AAY84538
 ID AAY84538 standard; protein; 1171 AA.

XX AC AAY84538;
 XX 25-JUL-2000 (first entry)

DE A chimeric collagen 1 (alpha1)/TGF-beta1 protein.

XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 KW trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 KW collagen; fibrinogen; fibronectin; post translational hydroxylation;
 KW ss. transforming growth factor-beta1; TGF-beta1; chimera.

XX Homo sapiens.
 OS Unidentified.
 OS Chimeric.

XX Key Location/Qualifiers
 FT Misc-difference 858
 FT /note= "Gly encoded by GCT"

XX EP992586-A2.

XX 12-APR-2000.

XX 07-OCT-1999; 99EP-00119184.

XX 09-OCT-1998; 98US-00169768.

XX (USSU) US SURGICAL CORP.

XX Gruskin EA, Buechter DD, Zhang G, Connolly K;

XX WPI; 2000-259138/23.

XX N-PSDB; AAA12498.

PT Production of extracellular matrix proteins containing 4-trans-
 PT hydroxyproline results in native self aggregating proteins, useful on
 PT medical implants.

XX Claim 23; Fig 15; 260pp; English.

XX The specification describes a method for producing an extracellular
 CC matrix protein or its fragment. The extracellular matrix protein is
 CC capable of self aggregating in a cell which does not ordinarily
 CC hydroxylated prolines. The method comprises optimising a nucleic acid
 CC sequence for expression in the cell by substitution of codons preferred
 CC by that cell for naturally occurring codons not preferred by the cell;
 CC incorporating the nucleic acid sequence into the cell; and contacting the
 CC cell with a hypertonic growth medium containing at least one amino acid,
 CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated

CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibrinogen and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents chimeric collagen 1 (alpha1)/transforming
 CC growth factor-beta1 (TGF-beta1) protein, which may be produced using the
 CC method of the invention

XX Sequence 1171 AA;

Query Match 90.5%; Score 38; DB 3; Length 1171;
 Best Local Similarity 100.0%; Pred. No. 9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEKSPG 7
 |||||
 Db 858 GAEKSPG 864

RESULT 56
 ADE87057
 ID ADE87057 standard; protein; 1211 AA.

XX AC ADE87057;

XX 29-JAN-2004 (first entry)

DE Human pancreatic cell protein sequence SeqID517.

XX neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;
 KW cancer death; cytostatic; vaccine; gene therapy;
 KW non-cancerous pancreas disease; human.

XX Homo sapiens.

XX WO2003060145-A2.

XX 24-JUL-2003.

XX 19-DEC-2002; 2002WO-US040655.

XX 21-DEC-2001; 2001US-0342768P.

XX (DIAD-) DIADEXUS INC.

XX Sun Y, Liu C;

XX WPI; 2003-587286/55.

XX N-PSDB; ADE87397.

PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
 PT staging, imaging, monitoring, preventing or treating pancreatic cancer or
 PT non-cancerous disease states of the pancreas.

XX Claim 12; SEQ ID NO 517; 635pp; English.

XX This invention relates to novel nucleic acids and proteins present in
 CC normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 CC cause of cancer death worldwide, therefore accurate methods of diagnosis
 CC and treatment are required. Compounds which modulate the proteins of the
 CC invention may have cytostatic activity and the protein and DNA sequences
 CC of the invention may be useful for the development of a vaccine or in
 CC gene therapy. The composition and methods are useful in diagnosing,
 CC staging, imaging, monitoring, preventing or treating pancreatic cancer
 CC and non-cancerous disease states of the pancreas. The present sequence is
 CC that of a human pancreatic protein of the invention.

XX Sequence 1211 AA;

Query Match 90.5%; Score 38; DB 7; Length 1211;
 Best Local Similarity 100.0%; Pred. No. 9.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 1019 GAEGSPG 1025

RESULT 57

ID ADE87062 standard; protein; 1226 AA.

XX AC ADE87062;

DT 29-JAN-2004 (first entry)

XX DE Human pancreatic cell protein sequence SeqID522.

XX KW neoplastic pancreatic cell; pancreatic cell; pancreatic cancer;

XX KW cancer death; cytostatic; vaccine; gene therapy;

XX KW non-cancerous pancreas disease; human.

XX OS Homo sapiens.

XX PN WO2003060145-A2.

XX PD 24-JUL-2003.

XX PF 19-DEC-2002; 2002WO-US040655.

XX PR 21-DEC-2001; 2001US-0342768P.

XX PA (DIAD-) DIADEXUS INC.

XX PI Sun Y, Liu C;

XX DR WPI; 2003-587286/55.

XX DR N-PSDB; ADE87403.

XX PT New pancreatic specific nucleic acid molecule or protein for diagnosing,
 staging, imaging, monitoring, preventing or treating pancreatic cancer or
 non-cancerous disease states of the pancreas.

XX PS Claim 12; SEQ ID NO 522; 635pp; English.

XX CC This invention relates to novel nucleic acids and proteins present in
 normal and neoplastic pancreatic cells. Pancreatic cancer is a common
 cause of cancer death worldwide, therefore accurate methods of diagnosis
 and treatment are required. Compounds which modulate the proteins of the
 invention may have cytostatic activity and the protein and DNA sequences
 of the invention may be useful for the development of a vaccine or in
 gene therapy. The composition and methods are useful in diagnosing,
 staging, imaging, monitoring, preventing or treating pancreatic cancer
 and non-cancerous disease states of the pancreas. The present sequence is
 that of a human pancreatic protein of the invention.

XX SQ Sequence 1226 AA;

Query Match 90.5%; Score 38; DB 7; Length 1226;
 Best Local Similarity 100.0%; Pred. No. 9.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 1019 GAEGSPG 1025

RESULT 58

ID ADS98184 standard; protein; 1284 AA.

XX AC ADS98184;

XX

DT 30-DEC-2004 (first entry)

XX DE Protein factor discovery related isolated human polypeptide, SEQ ID 448.
 XX KW antiinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;
 KW leukaemia; nervous system disorder; infection.

XX OS Homo sapiens.

XX PN WO2004087874-A2.

XX PD 14-OCT-2004.

XX PF 24-MAR-2004; 2004WO-US009202.

XX PR 28-MAR-2003; 2003US-0458824P.

XX PA (NUVE-) NUVELO INC.

XX PA (DRMA/) DRMANAC R T.

XX PI Tang YT, Zhou P, Wang J, Wang ZW, Hu T;

XX DR WPI; 2004-737686/72.

XX DR N-PSDB; ADS97949.

XX PT New polynucleotides encoding a polypeptide with biological activity,
 useful for treating inflammation, leukaemia, nervous system disorders, or
 PT infections.

XX PS Claim 20; SEQ ID NO 448; 253pp; English.

XX CC The invention relates to a novel isolated polynucleotide comprising any
 of the 235 nucleotide sequences described in the specification. The
 invention further comprises: an isolated polynucleotide encoding a
 polypeptide with biological activity, where the polynucleotide hybridizes
 to one of the 235 novel polynucleotides under stringent hybridization
 conditions, or having greater than about 99% sequence identity with the
 novel polynucleotide; a vector comprising a novel polynucleotide; an
 expression vector comprising the novel polynucleotide; a host cell
 genetically engineered to comprise the novel polynucleotide, which can be
 operatively associated with a regulatory sequence that modulates
 expression of the polynucleotide in the host cell; an isolated
 polypeptide encoded by the novel polynucleotide, or a polynucleotide
 hybridizing under stringent conditions to the novel polynucleotide; a
 composition comprising the polypeptide and a carrier; an antibody
 directed against the polypeptide; a method for detecting the novel
 polynucleotide in a sample; a method for detecting the polypeptide in a
 sample; a method for identifying a compound that binds to the polypeptide
 ; a method for producing the polypeptide; an isolated polypeptide
 comprising any of the 235 amino acid sequences described in the
 specification; and a collection of polynucleotides comprising of at least
 one of the polynucleotides cited above. The polypeptides and
 polynucleotides of the invention have antiinflammatory, cytostatic, and
 antimicrobial activities. The novel polynucleotide may be used to treat
 disorders by gene therapy. The polypeptides and polynucleotides are
 useful for treating inflammation, leukaemia, nervous system disorders,
 or infections. This sequence represents one of the 235 novel isolated
 polypeptides of the invention.

XX SQ Sequence 1284 AA;

Query Match 90.5%; Score 38; DB 8; Length 1284;
 Best Local Similarity 100.0%; Pred. No. 9.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||

Db 839 GAEGSPG 845

RESULT 59

AAR71701

```

ID AAR71701 standard; protein; 1341 AA.
XX
AC AAR71701;
XX
DT 25-MAR-2003 (revised)
DT 17-OCT-1995 (first entry)
XX
XX Collagen alpha 1 (I) chain precursor.
DE
XX Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
KW disorder; osteoporosis; metastatic progression; Paget's disease;
KW hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis;
KW vasculitis syndrome.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 2028
FT /note= "Unidentified amino acid."
XX
XX WO9508115-A1.
XX
XX 23-MAR-1995.
XX
XX 19-SEP-1994; 94WO-DK000348.
XX
XX 17-SEP-1993; 93DK-00001040.
XX
XX (OSTE-) OSTEOMETER AS.
XX
XX Qvist P, Bonde M;
XX
XX WPI; 1995-131456/17.
XX
XX Assaying collagen fragments in body fluid by immunoassay - using
PT antibodies raised against synthetic peptide(s) contg. potential
PT crosslinking sites, to diagnose and monitor disorders of collagen
PT metabolism, e.g. osteoporosis.
XX
XX Disclosure (Appendix A); Page 49; 87pp; English.
XX
XX Determination of collagen fragments in body fluids can be achieved by
CC immunoassay using antibodies directed against synthetic peptides derived
CC from collagen which contain sites of potential crosslinking. The method
CC is used to diagnose and monitor treatment of disorders of collagen
CC metabolism (degradation of type I collagen may indicate osteoporosis,
CC metastatic progression, Paget's disease, hyperthyroidism or other
CC conditions involving excessive bone resorption; degradation of type II
CC collagen may indicate rheumatoid arthritis or osteoarthritis; and of type
CC III collagen, vaculitis syndrome). The method can also be used to assess
CC the toxicity of a compound and to test drugs for their effect on collagen
CC metabolism. (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 1341 AA;
Query Match 90.5%; Score 38; DB 2; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 896 GAEGSPG 902

RESULT 60
AAY96122
ID AAY96122 standard; peptide; 1341 AA.
XX
AC AAY96122;
XX
XX 19-DEC-2000 (first entry)
DT
XX
XX Collagen type I alpha-1.
DE

XX
KW Collagen type I; osteoporosis; bone resorption; Paget's disease;
KW hyperparathyroidism; metastasis; assay; diagnosis.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 924
FT /note= "unidentified residue"
XX
XX Misc-difference 927
FT /note= "unidentified residue"
XX
XX Misc-difference 1127
FT /note= "unidentified residue"
XX
XX Misc-difference 1268
FT /note= "unidentified residue"
XX
XX US6110689-A.
PN
XX 29-AUG-2000.
XX
XX 04-NOV-1997; 97US-00963825.
XX
XX 21-JAN-1994; 94US-00187319.
XX
XX (OSTE-) OSTEOMETER AS.
XX
XX Bonde M, Qvist P;
XX
XX WPI; 2000-586349/55.
XX
XX Assaying type I collagen fragments for diagnosing osteoporosis in
PT postmenopausal woman, involves contacting body fluid with synthetic
PT collagen peptide and antibody and quantifying by competitive binding
PT assay.
XX
XX Disclosure; Col 23-37; 41pp; English.
XX
XX The present sequence is that of human type I collagen alpha-1. The
CC invention is based on the discovery of the presence of particular
CC collagen fragments in body fluids of patients compared with those of
CC healthy subjects. These fragments are generated upon collagen degradation
CC and are partly characterised by the presence of potential sites for
CC crosslinking. A method for assaying collagen fragments in a body fluid
CC sample is based on the competitive binding to immunological binding
CC partners of collagen fragments in the sample and of synthetic peptides
CC derived from collagen and containing crosslinkable sites (see AAY96105-
CC 11). When considering the degradation of type I collagen, the assay can
CC be used as a means of identifying excessive bone resorption, indicating
CC the presence of osteoporosis or the metastatic progress of a malignancy.
CC Other conditions characterized by excessive bone resorption include
CC Paget's disease and hyperparathyroidism
XX
XX Sequence 1341 AA;
Query Match 90.5%; Score 38; DB 3; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 895 GAEGSPG 901

RESULT 61
AAE16475
ID AAE16475 standard; protein; 1341 AA.
XX
AC AAE16475;
XX
XX 09-APR-2002 (first entry)
DT
XX
XX Human collagen alphas (I) protein.
DE

```


KW Human; collagen; osteoporosis; bone resorption; Paget's disease;
 KW hyperparathyroidism; rheumatoid arthritis; osteoarthritis; therapy;
 KW vasculitis syndrome; toxic substance; tissue degradation; algal (I).
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Key 924 /label= Unknown
 XX FT Misc-difference 927 /label= Unknown
 XX FT Misc-difference 1127 /label= Unknown
 XX FT Misc-difference 1268 /label= Unknown
 XX FT Misc-difference 1268 /label= Unknown
 XX
 XX US6323314-B1.
 XX
 XX 27-NOV-2001.
 XX
 XX 10-FEB-2000; 2000US-00500811.
 XX
 XX 21-JAN-1994; 94US-00187319.
 XX 04-NOV-1997; 97US-00963825.
 XX
 XX (OSTE-) OSTEOMETER AS.
 XX
 XX Qvist P, Bonde M;
 XX WPI; 2002-096598/13.
 XX
 XX Novel peptides that match alpha 1 or 2(I) telopeptide component of cross-linked telopeptide degradation product of type I collagen useful for diagnosing disorders associated with collagen metabolism e.g. osteoporosis.
 XX
 XX Disclosure; Col 23-30; 34pp; English.
 XX
 XX The invention relates to a peptide synthesised to match an alpha(I) or alpha2(I) telopeptide component of a cross-linked telopeptide degradation product of type I collagen. The method is useful for assaying collagen fragments in animal body fluids, for determining the degradation of human collagen types I, II and III and for diagnosing the presence of disorders associated with the metabolism of collagen, especially osteoporosis. The peptide is used for assessing the impact of drugs on collagen metabolism. The peptide is useful in methods to assess an abnormal condition of a subject for e.g. excessive bone resorption which shows the presence of an osteoporotic condition or the metastatic progress of a malignancy, Paget's disease and hyperparathyroidism. Disease states involving connective tissues can be monitored by determining collagen degradation, examples are collagen type II degradation associated with rheumatoid arthritis, osteoarthritis, and collagen type III degradation in vasculitis syndrome. Since the conditions of the subject can be monitored continuously, application of these assays can also be used to monitor the progress of therapy administered to treat these or other conditions and as a measure of toxicity, since the administration of toxic substances often results in tissue degradation. The present sequence is human collagen alpha1 (I) protein used in the invention
 XX Sequence 1341 AA;
 XX
 XX Query Match 90.5%; Score 38; DB 5; Length 1341;
 XX Best Local Similarity 100.0%; Pred. No. 1e+03;
 XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1 GAEGSPG 7
 XX | | | | |
 XX 895 GAEGSPG 901
 XX
 XX DB 895 GAEGSPG 901
 XX
 XX RESULT 62
 XX ABB0733
 XX ID ABB0733 standard; protein; 1341 AA.

XX ABB0733;
 XX
 XX 15-JUL-2002 (first entry)
 XX
 XX Collagen type I-alpha1 protein.
 XX
 XX Collagen; osteoarthritis; Paget's disease; Marfan syndrome; dwarfism;
 XX osteogenesis imperfecta; neoplastic growth; rheumatoid arthritis;
 XX vasculitis; collagen type I-alpha1.
 XX
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX FT Misc-difference 1. 1341
 XX FT /note= "residues Xaa are unknown"
 XX
 XX US6355442-B1.
 XX
 XX 12-MAR-2002.
 XX
 XX 13-APR-2000; 2000US-00548608.
 XX
 XX 21-JAN-1994; 94US-00187319.
 XX 04-NOV-1997; 97US-00963825.
 XX
 XX (OSTE-) OSTEOMETER BIOTECH AS.
 XX
 XX Qvist P, Bonde M;
 XX WPI; 2002-380937/41.
 XX
 XX Assaying type I collagen fragments in body fluid, useful for diagnosis and assessing treatment of e.g. osteoarthritis, by competitive immunoassay.
 XX
 XX Disclosure; Col 23-30; 35pp; English.
 XX
 XX The invention relates to a method for assaying type I collagen fragments (I) in body fluid. The method involves treating the test sample with: (i) synthetic peptide, immobilised on a support; and (ii) immunological binding partner, reactive with the synthetic peptide, so that (i) and the synthetic peptide compete for binding, and (i) are quantified by measuring the binding of the binding partner to the synthetic peptide. The method is used to diagnose disorders of collagen metabolism, especially osteoarthritis but also Paget's disease, Marfan syndrome, osteogenesis imperfecta, neoplastic growth of collagenous tissue, dwarfism, rheumatoid arthritis or vasculitis, also for clinical testing of drugs to assess their effect on collagen metabolism. The present sequence represents the collagen type I-alpha1 protein
 XX Sequence 1341 AA;
 XX
 XX Query Match 90.5%; Score 38; DB 5; Length 1341;
 XX Best Local Similarity 100.0%; Pred. No. 1e+03;
 XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX QY 1 GAEGSPG 7
 XX | | | | |
 XX 895 GAEGSPG 901
 XX
 XX DB 895 GAEGSPG 901
 XX
 XX RESULT 63
 XX ABB09625
 XX ID ABB09625 standard; peptide; 1341 AA.
 XX
 XX AC ABB09625;
 XX
 XX 29-MAY-2002 (first entry)
 XX
 XX Amino acid sequence of human collagen type I alpha1.
 XX
 XX Collagen; bone resorption; collagen metabolism; Paget's disease;

KW Marfan's syndrome; osteogenesis imperfecta; neoplastic growth; dwarfism;
KW rheumatoid arthritis; osteoarthritis; vasculitis syndrome.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 924 /note= "unspecified amino acid"
FT Misc-difference 927 /note= "unspecified amino acid"
FT Misc-difference 1127 /note= "unspecified amino acid"
FT Misc-difference 1268 /note= "unspecified amino acid"
XX
XX US6342361-B1.
XX
XX 29-JAN-2002.
XX
XX 12-MAY-2000; 2000US-00570573.
XX
XX 21-JAN-1994; 94US-00187319.
PR 04-NOV-1997; 97US-00963825.
XX
XX (OSTE-) OSTROMETER BIOTECH AS.
XX
XX Qvist P, Bonde M;
XX
XX WPI; 2002-224940/28.
XX
XX Assaying type II collagen fragments in a body fluid sample, for
PT determining abnormalities in collagen metabolism, e.g. in Paget's disease
PT or Marfan's syndrome, comprises contacting the fluid with a synthetic
PT type II collagen alpha.
XX
XX Disclosure; Col 23-30; 35pp; English.
XX
XX The present sequence represents human collagen type I alpha1. The
CC specification describes a method for assaying type II collagen fragments
CC in a body fluid sample. The method comprises contacting the body fluid
CC with a synthetic peptide consisting essentially of a type II collagen
CC alpha1 amino acid sequence that competes with the collagen fragments for
CC binding with the immunological binding partner. The method is useful for
CC assaying collagen fragments in body fluids. The method is particularly
CC useful in assays for measuring bone resorption rates in patients or in
CC healthy subjects. The method may be used for determining abnormalities in
CC collagen metabolism, especially in Paget's disease, Marfan's syndrome,
CC osteogenesis imperfecta, neoplastic growth in collagenous tissue,
CC dwarfism, rheumatoid arthritis, osteoarthritis or vasculitis syndrome.
CC The method may be used for determining the degradation of human collagen
CC of type I, II and III. It also can be used during clinical testing of new
CC drugs to assess the impact of these drugs on collagen metabolism. The
CC assays, can be used as a measure of toxicity, since the administration of
CC toxic substances often results in tissue degradation. Thus, the assays
CC may be applied in any situation where the metabolic condition of collagen
CC tissues can be used as an index of the condition, treatment or effect of
CC substances directly administered to the subject or to which the subject
CC is exposed in the environment
XX
XX Sequence 1341 AA;
SQ
Query Match 90.5%; Score 38; DB 5; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
Db 895 GAEGSPG 901
|||||
895 GAEGSPG 901
RESULT 64
ADFL13075
ID ADF13075 standard; protein; 1341 AA.

XX ADF13075;
XX
XX 12-FEB-2004 (first entry)
XX
XX Human collagen alpha(I) chain precursor.
DE
XX collagen fragment; collagen; collagen metabolism disorder;
KW collagen degradation; crosslinking site; human; alpha1(I) chain.
KW
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 924 /note= "Unknown"
FT Misc-difference 927 /note= "Unknown"
FT Misc-difference 1127 /note= "Unknown"
FT Misc-difference 1268 /note= "Unknown"
XX
XX US2003119058-A1.
PN
XX 26-JUN-2003.
XX
XX 29-JAN-2002; 2002US-00058124.
XX
XX 21-JAN-1994; 94US-00187319.
PR 04-NOV-1997; 97US-00963825.
PR 12-MAY-2000; 2000US-00570573.
XX
XX (OSTE-) OSTROMETER AS.
PA
XX Qvist P, Bonde M;
XX
XX WPI; 2003-897106/82.
XX
XX Determining collagen fragments in a body fluid sample using antibodies
PT against synthetic peptides containing crosslinking sites is useful to
PT diagnose collagen metabolism disorders or if a subject suffers from
PT collagen degradation.
XX
XX Example 3; SEQ ID NO 18; 41pp; English.
XX
XX The invention relates to a new method for determining collagen fragments
CC in a body fluid which comprises contacting the sample with an
CC immunological binding partner for the fragments, where the binding
CC partner is immunoreactive with synthetic peptides whose sequences are
CC essentially derived from collagen and contain potential sites for cross
CC linking, and is incorporated as a whole antibody or its immunological
CC fragment in an assay for quantitative determination of collagen fragments
CC in the sample. The invention is useful for diagnosing disorders
CC associated with the metabolism of collagen or whether a susceptible
CC subject is suffering from collagen degradation. A test kit for
CC quantitating the amount of collagen fragments in a body fluid is claimed.
CC The present sequence represents the amino acid sequence of human collagen
CC alpha1(I) chain precursor.
XX
XX Sequence 1341 AA;
SQ
Query Match 90.5%; Score 38; DB 7; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
Db 895 GAEGSPG 901
|||||
895 GAEGSPG 901
RESULT 65
AAY84539
ID AAY84539 standard; protein; 1388 AA.

XX AAY84539;
 XX 25-JUL-2000 (first entry)
 XX Amino acid sequence of a chimeric collagen 1 (alpha1)/decorin protein.
 XX Extracellular matrix protein; self aggregation; hydroxylated proline;
 XX trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production;
 XX collagen; fibronectin; fibronectin; post translational hydroxylation;
 XX decorin; chimera.
 XX Homo sapiens.
 XX Unidentified.
 XX Chimeric.
 XX Key Location/Qualifiers
 XX Misc-difference 87 /note= "Gly encoded by GCG"
 XX Misc-difference 305 /note= "Glu encoded by CAA"
 XX Misc-difference 363 /note= "Gly encoded by CGT"
 XX Misc-difference 378 /note= "Glu encoded by GGT"
 XX Misc-difference 429 /note= "Gly encoded by CGA"
 XX Misc-difference 444 /note= "Gly encoded by GCG"
 XX Misc-difference 543 /note= "Gly encoded by GCC"
 XX Misc-difference 546 /note= "Gly encoded by GCT"
 XX Misc-difference 606 /note= "Gly encoded by GAC"
 XX Misc-difference 702 /note= "Gly encoded by GGT"
 XX Misc-difference 815 /note= "Pro encoded by CTT"
 XX Misc-difference 858 /note= "Gly encoded by GCT"
 XX Misc-difference 1066 /note= "Gly encoded by GCC"
 XX EP992586-A2.
 XX 12-APR-2000.
 XX 07-OCT-1999; 99EP-00119184.
 XX 09-OCT-1998; 98US-00169768.
 XX (USSU) US SURGICAL CORP.
 XX Gruskin EA, Buechter DD, Zhang G, Connolly K;
 XX WPI; 2000-259138/23.
 XX N-PSDB; AAA12499.
 XX Production of extracellular matrix proteins containing 4-trans-
 XX hydroxyproline results in native self aggregating proteins, useful on
 XX medical implants.
 XX Claim 25; Fig 17A-B; 260pp; English.
 XX The specification describes a method for producing an extracellular
 XX matrix protein or its fragment. The extracellular matrix protein is
 XX capable of self aggregating in a cell which does not ordinarily
 XX hydroxylated prolines. The method comprises optimising a nucleic acid
 XX sequence for expression in the cell by substitution of codons preferred
 XX by that cell for naturally occurring codons not preferred by the cell;
 XX incorporating the nucleic acid sequence into the cell; and contacting the
 XX cell with a hypertonic growth medium containing at least one amino acid,

CC selected from the group consisting of trans-4-hydroxyproline and 3-
 CC hydroxyproline to allow at least one of the amino acids to be assimilated
 CC into the cell and incorporated into the extracellular matrix protein. The
 CC method may be used to make host cells assimilate and incorporate trans-4-
 CC hydroxyproline into proteins. This is especially useful in the
 CC recombinant production of proteins such as collagen, fibronectin and
 CC fibronectin whose ability to self aggregate and produce functional
 CC proteins depends on the post translational hydroxylation of proline. The
 CC method is also useful in studying the structure and function of
 CC polypeptides which do not normally contain trans-4-hydroxyproline. The
 CC present sequence represents a chimeric collagen 1 (alpha1)/decorin
 CC protein, which may be produced using the method of the invention
 XX
 XX SQ Sequence 1388 AA;
 Query Match 90.5%; Score 38; DB 3; Length 1388;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAGGSPG 7
 Db 858 GAGGSPG 864
 RESULT 66
 AAY56800
 ID AAY56800 standard; protein; 1411 AA.
 XX
 AC AAY56800;
 XX 27-MAR-2000 (first entry)
 XX Human preproalpha 1 (I) collagen.
 XX Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin;
 XX sequence selection and alignment domain; prosthetic implant; foodstuff;
 XX medicine; type I collagen; human.
 XX Homo sapiens.
 XX EP967226-A2.
 XX 29-DEC-1999.
 XX 04-MAY-1999; 99EP-00303470.
 XX 08-MAY-1998; 98US-0084828P.
 XX 10-APR-1999; 99US-00289578.
 XX (COHE-) COHESION TECHNOLOGIES INC.
 XX Olsen DR, Hitzeman RA, Chisholm GE;
 XX WPI; 2000-074666/07.
 XX New method for production of fibrillar collagen, useful for preparing
 XX telopeptide collagen fibrils and gelatin.
 XX Example 1; Fig 3A-B; 30pp; English.
 XX The invention provides a method for the production of fibrillar collagen.
 XX The method comprises: (a) culturing a recombinant host cell comprising a
 XX DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD
 XX (sequence selection and alignment domain); and (b) producing the
 XX fibrillar collagen. The methods are used to produce fibrillar collagen,
 XX from which telopeptide collagen fibrils can be derived. Host cells,
 XX comprising DNA encoding a collagen monomer lacking SSAD or N propeptide
 XX is used to produce gelatin. Collagen is used in biological research as a
 XX substrate for in vitro cell culture and as a component of biocompatible
 XX materials for use in prosthetic implants, sustained drug release
 XX matrices, artificial skin and wound dressing and healing devices. Gelatin
 XX is particularly useful for foodstuffs and medicine, for coating tablets
 XX and making capsules. The methods, comprising the use of collagen monomers

CC lacking the N and/or C propeptides, result in a large increase in the
CC production of type I collagen. The present sequence represents the human
CC preproalpha 1 (I) collagen (GenBank Accn no: AF017178)

XX
SQ Sequence 1411 AA;

Query Match 90.5%; Score 38; DB 3; Length 1411;

Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 1 GAEGSPG 7

Db 1016 GAEGSPG 1022
|||||||

RESULT 67

AAE02535
ID AAE02535 standard; protein; 1449 AA.

XX AC AAE02535;

XX DT 10-AUG-2001 (first entry)

XX DE Porcine alpha1(I) collagen.

XX KW Porcine; alpha1(I) collagen; gelatin; cytostatic; viral infection;
XX KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
XX KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
XX KW rheumatoid arthritis; beverage; photographic application.

XX OS Sus scrofa.

XX PH Key Location/Qualifiers

FT Misc-difference 829. .830

FT FT /note= "Encoded by ggcgaaacctggtgatgctgctgctaaaggcgat
ctggtcccccggcctgtgga"

XX PN WO200134647-A2.

XX PD 17-MAY-2001.

XX PF 10-NOV-2000; 2000MO-US030792.

XX PR 12-NOV-1999; 99US-00439058.

XX PR 10-NOV-2000; 2000US-00709700.

XX PA (FIBR-) FIBROGEN INC.

XX PI Bell MP, Neff TB, Polarek JW, Seeley TW;

XX DR WPI; 2001-335911/35.

XX DR N-PSDB; AAD06576.

XX PT Novel isolated and purified bovine or porcine collagens and gelatins
XX PT useful in medical, pharmaceutical, food and cosmetic industries, as
XX PT vaccine, and for treating autoimmune disorders, infections and cancer.

XX PS Example 3; Fig 8; 168pp; English.

XX CC The present sequence is porcine alpha1(I) collagen. The present invention
XX CC relates to recombinant synthesis of collagens and gelatins derived from
XX CC animals. Collagen is useful in medical, pharmaceutical, food and cosmetic
XX CC industries. Collagen is an important component of arterial sealants, bone
XX CC grafts, drug delivery system, dermal implants, haemostats, and
XX CC incontinence implants, and for treating autoimmune disorders such as
XX CC rheumatoid arthritis. Collagen is useful in food products such as sausage
XX CC casings, and in cosmetics or facial and skin products such as
XX CC moisturisers. Recombinant gelatin is useful in vaccine formulations for
XX CC treating viral infections, autoimmune diseases and cancer. Gelatin is
XX CC useful in the manufacture or as a component of various pharmaceutical and
XX CC medical devices and products, in food and beverage industries, in hair
XX CC care and skin care products, as a glue or adhesive in various
XX CC manufacturing processes, as a light-sensitive coating in various

CC electronic devices, as photoresist base in photolithographic processes,
CC in printing and photographic applications, in laboratory application, and
CC as a component in various gels used for biochemical and electrophoretic
CC analysis, including enzymographic gels

XX SQ Sequence 1449 AA;

Query Match 90.5%; Score 38; DB 4; Length 1449;

Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

OY 1 GAEGSPG 7

Db 1004 GAEGSPG 1010
|||||||

RESULT 68

ABG93948
ID ABG93948 standard; protein; 1453 AA.

XX AC ABG93948;

XX DT 26-NOV-2002 (first entry)

XX DE Mouse polypeptide orthologous to DACC-11.

XX KW Human; deer; rat; mouse; DACC; deer antler cartilage cell;
XX KW cell stimulation; cell inhibition; cell growth; cell division;
XX KW mesenchymal cell; chondrocyte; chondrogenesis; osteogenesis; growth;
XX KW repair; regeneration; restoration; extracellular matrix;
XX KW cartilaginous matrix; cartilage; disc; connective tissue; agonist;
XX KW antagonist; gene therapy.

XX OS Mus sp.

XX PN WO200264625-A1.

XX PD 22-AUG-2002.

XX PF 15-FEB-2002; 2002WO-AU000163.

XX PR 15-FEB-2001; 2001AU-00003116.

XX PA (ADPP-) ADP PHARM PTY LTD.

XX PA (UNSY) UNIV SYDNEY.

XX PI Roubin R, Ghosh P;

XX DR WPI; 2002-643456/69.

XX PT Stimulating or inhibiting cell growth and/or division, useful for
XX PT stimulating chondrogenesis, cartilage, disc or connective tissue growth,
XX PT repair, and/or regeneration, comprises administering deer antler
XX PT cartilage gene.

XX PS Claim 13; Page 160-166; 214pp; English.

XX CC The invention discloses a method for stimulating or inhibiting cell
XX CC growth and/or division which comprises contacting or inserting into an
XX CC animal cell a polypeptide comprising one of the deer antler cartilage
XX CC cell (DACC) clones disclosed. More particularly, the method relates to
XX CC these polypeptides stimulating mesenchymal cell growth and/or division
XX CC and to transfecting these cells and chondrocytes with vectors carrying
XX CC the genes of these polypeptides capable of stimulating chondrogenesis,
XX CC osteogenesis, growth, repair, regeneration and/or restoration of the
XX CC extracellular matrix. The chondrocytes selectively express genes required
XX CC to form a cartilaginous matrix. The DACC polypeptides and polynucleotides
XX CC are useful for identifying an agent that modulates the activity of the
XX CC polypeptide, for stimulating mesenchymal cell growth and/or division by
XX CC exposing animal mesenchymal cells to conditioned media or its active
XX CC fraction, obtained from deer antler cartilage cells, for inhibiting cell
XX CC growth and/or division by inserting into an animal cell, a compound which
XX CC inhibits the translation of the polynucleotide encoding the DACC. The

CC method and the polypeptides are useful for stimulating mesenchymal cell
 CC growth and/or division or for stimulating chondrogenesis, cartilage, disc
 CC or connective tissue growth, repair, regeneration and/or restoration in
 CC an animal. The polynucleotides, polypeptides, agonists and antagonists
 CC may be used in treatment modalities, specifically in gene therapy. The
 CC polypeptides can be used as bait proteins in a two- or three-hybrid assay
 CC to identify other proteins, which bind to or interact with the
 CC polypeptide and are involved in modulating cell growth and/or division.
 CC The sequences presented in ABG93923-ABG93948 are the proteins encoded by
 CC the DACC cDNA clones
 XX

SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 5; Length 1453;

Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0;

Qy 1 GAEGSPG 7
 |||||

Db 1008 GAEGSPG 1014

RESULT 69

ADD45053
 ID ADD45053 standard; protein; 1453 AA.

XX

AC ADD45053;

XX

DT 29-JAN-2004 (first entry)

XX

DE Rat Protein CAB01633, SEQ ID NO 10485.

XX

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX

OS Rattus norvegicus.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR

PR 01-NOV-2001; 2001US-0346382P.

XX

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO) GEN HOSPITAL CORP.

XX

PA (FARB) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

PI WPI; 2003-268312/26.

XX

DR GENBANK; CAB01633.

XX

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 7; Length 1453;

Best Local Similarity 100.0%; Pred. No. 1.1e+03; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0;

Qy 1 GAEGSPG 7
 |||||

Db 1008 GAEGSPG 1014

RESULT 70

ADD45057

ID ADD45057 standard; protein; 1453 AA.

XX

AC ADD45057;

XX

DT 29-JAN-2004 (first entry)

XX

DE Rat Protein CAB01633, SEQ ID NO 10489.

XX

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX

OS Rattus norvegicus.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR

PR 01-NOV-2001; 2001US-0346382P.

XX

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO) GEN HOSPITAL CORP.

XX

PA (FARB) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

PI WPI; 2003-268312/26.

XX

DR GENBANK; CAB01633.

XX

PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 7; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
 |||||
 Db 1008 GAEGSPG 1014

RESULT 71
 ADD48341
 ID ADD48341 standard; protein; 1453 AA.

AC ADD48341;
 XX
 DT 02-DEC-2004 (revised)
 DT 29-JAN-2004 (first entry)
 XX
 XX Rat Protein CAB01633, SEQ ID NO 14041.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 KW
 XX Rattus norvegicus.
 OS Unidentified.

XX WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 XX 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX

XX (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 PA

XX Woolf C, D'urso D, Belfort K, Costigan M;
 XX WPI; 2003-268312/26.
 DR GENBANK; CAB01633.

XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 PT

XX Example 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (described in Table 3
 CC of the specification) which is differentially expressed during pain.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 7; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
 |||||
 Db 1008 GAEGSPG 1014

RESULT 72
 ADD45049
 ID ADD45049 standard; protein; 1453 AA.

AC ADD45049;
 XX

XX 29-JAN-2004 (first entry)
 DT

XX Rat Protein CAB01633, SEQ ID NO 10481.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 KW

XX Rattus norvegicus.
 OS

XX WO2003016475-A2.
 PN

XX 27-FEB-2003.
 PD

XX 14-AUG-2002; 2002WO-US025765.
 XX

XX 14-AUG-2001; 2001US-0312147P.
 PR

XX 01-NOV-2001; 2001US-0346382P.
 PR

XX 26-NOV-2001; 2001US-0333347P.
 PR

XX (GEHO) GEN HOSPITAL CORP.
 PA

PA (FARB) BAYER AG.
 PA

XX Woolf C, D'urso D, Belfort K, Costigan M;
 PI

XX WPI; 2003-268312/26.
 DR GENBANK; CAB01633.

XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 PT

XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (shown in Table 2 of
 CC the specification) which is differentially expressed during pain. Note:
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 7; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 1008 GAEGSPG 1014

RESULT 73
 ADD48337
 ID ADD48337 standard; protein; 1453 AA.

XX AC ADD48337;
 XX 02-DEC-2004 (revised)
 DT 29-JAN-2004 (first entry)
 XX Rat Protein CAB01633, SEQ ID NO 14037.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX Rattus norvegicus.
 OS Unidentified.
 XX WO2003016475-A2.

XX 27-FEB-2003.
 XX 14-AUG-2002; 2002WO-US025765.
 XX 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.

XX (GHEO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;
 XX WPI; 2003-268312/26.

DR GENBANK; CAB01633.

XX New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.

XX Example 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating
 CC pain and a pharmaceutical composition comprising the one or more
 CC polypeptides or their antibodies. The polynucleotide or the compound that
 CC modulates its activity is useful for preparing a medicament for treating
 CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
 CC therapy). The sequence presented is a rat protein (described in Table 3
 CC of the specification) which is differentially expressed during pain.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic form directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 1453 AA;

Query Match 90.5%; Score 38; DB 7; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 1.1e+03;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
 |||||
 Db 1008 GAEGSPG 1014

RESULT 74
 ADD48345
 ID ADD48345 standard; protein; 1453 AA.

XX AC ADD48345;

XX 02-DEC-2004 (revised)
 DT 29-JAN-2004 (first entry)

XX Rat Protein CAB01633, SEQ ID NO 14045.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX Rattus norvegicus.
 OS Unidentified.

XX WO2003016475-A2.

XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.

XX 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:01:18 ; Search time 18.8 Seconds
(without alignments)
40.943 Million cell updates/sec

Title: US-10-698-121A-1

Perfect score: 42

Sequence: 1 GAEGSPGL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

PIR 80:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	633	2 B40983	collagen alpha 1(X
2	39	92.9	1707	2 A33526	collagen alpha 2(I
3	38	90.5	198	2 I49558	collagen alpha 1(I
4	38	90.5	779	1 CG801S	collagen alpha 1(I
5	38	90.5	1453	2 S21626	collagen alpha 1(I
6	38	90.5	1464	1 CGHU1S	collagen alpha 1(I
7	37	88.1	812	2 S31521	collagen COLF1 - f
8	37	88.1	1027	2 S28774	collagen alpha cha
9	37	88.1	1414	1 S23809	collagen alpha 2(I
10	37	88.1	1464	2 S59856	collagen alpha 1(I
11	37	88.1	1690	1 CGHU1B	collagen alpha 4(I
12	36	85.7	486	2 A41537	DNA-binding protei
13	36	85.7	619	2 T08613	hypothetical prote
14	36	85.7	650	2 T22002	hypothetical prote
15	36	85.7	749	2 T40812	hypothetical glyci
16	36	85.7	1024	2 S18251	collagen alpha 1(X
17	36	85.7	1603	2 S23810	collagen alpha 1(X
18	36	85.7	1806	1 CGHU1E	collagen alpha 1(X
19	35	83.3	258	2 B89773	collagen col-34 -
20	35	83.3	298	2 JC1448	hypothetical prote
21	35	83.3	299	2 T29956	hypothetical prote
22	35	83.3	314	2 T32247	collagen alpha 1(I
23	35	83.3	325	2 S02170	hypothetical prote
24	35	83.3	358	2 T26281	collagen - chicken
25	35	83.3	473	2 I50629	collagen alpha 1(I
26	35	83.3	636	2 S41067	collagen alpha 3(I
27	35	83.3	675	2 S20819	collagen alpha 3(I
28	35	83.3	825	1 EDBEXD	immediate-early pr
29	35	83.3	888	2 S28791	collagen alpha 1(X

30	83.3	921	2	S40495	collagen alpha 1(I
31	83.3	921	2	S42617	collagen alpha 1(I
32	83.3	1042	1	CGCH1S	collagen alpha 1(I
33	83.3	1049	1	CGBO7S	collagen alpha 1(I
34	83.3	1306	2	A70934	hypothetical glyci
35	83.3	1466	1	CGHU7L	collagen alpha 1(I
36	83.3	1670	1	CGHU3B	collagen alpha 3(I
37	83.3	1752	2	A45407	collagen alpha 3(I
38	83.3	2274	2	T30258	adenomatous polyo
39	83.3	206	2	A75508	hypothetical prote
40	81.0	295	1	S42627	glycine N-methyltr
41	81.0	379	2	S55900	DNAJ-like protein
42	81.0	379	2	T41633	psi protein - fies
43	81.0	513	2	T31115	histidine kinase h
44	81.0	618	2	S32436	collagen alpha 2(I
45	81.0	754	2	A55267	collagen alpha 5(I
46	81.0	771	2	S72526	inorganic diphosph
47	81.0	1418	2	T45467	collagen alpha 1(I
48	81.0	1419	2	A41182	collagen alpha 1(I
49	81.0	1486	1	B40333	collagen alpha 1(I
50	81.0	1487	1	CGHU6C	collagen alpha 1(I
51	81.0	1487	2	B41182	collagen alpha 1(I
52	81.0	1492	2	A40333	collagen alpha 1(I
53	81.0	1549	2	I48103	collagen alpha 1(I
54	81.0	1669	1	CGHU4B	type VII collagen
55	81.0	1669	1	CGMS4B	collagen alpha 1(I
56	81.0	1691	1	S22917	collagen alpha 5(I
57	81.0	1747	2	A54121	collagen alpha-4 c
58	81.0	1763	2	S16366	collagen alpha 2(I
59	81.0	1775	2	A31893	collagen alpha 1(I
60	81.0	2944	2	A54849	collagen alpha 1(I
61	81.0	129	2	B83173	hypothetical prote
62	81.0	129	2	T24064	hypothetical prote
63	81.0	152	2	T26125	hypothetical prote
64	81.0	290	2	A32249	collagen - sea urc
65	81.0	290	2	B88638	protein F58F6.2 (i
66	81.0	297	2	T18637	hypothetical prote
67	81.0	310	2	I50696	collagen alpha 1(I
68	81.0	319	2	S76227	hypothetical prote
69	81.0	321	2	T28760	hypothetical prote
70	81.0	323	2	A61396	collagen alpha 1(I
71	81.0	327	2	T34203	hypothetical prote
72	81.0	359	2	T22774	hypothetical prote
73	81.0	366	2	S14449	collagen short cha
74	81.0	369	2	F96788	protein T4012.22 l
75	81.0	381	2	T27806	hypothetical prote
76	81.0	423	2	A41207	collagen I3, nonfi
77	81.0	458	2	T31631	hypothetical prote
78	81.0	482	2	B31795	collagen alpha 1(X
79	81.0	630	2	T48369	hypothetical prote
80	81.0	632	2	S42731	collagen alpha 1 c
81	81.0	671	1	CGRT1S	collagen alpha 1(I
82	81.0	680	2	S31216	collagen alpha 1(X
83	81.0	730	2	A36226	collagen alpha 1(I
84	81.0	886	2	I50694	collagen alpha 2(I
85	81.0	917	2	S09646	collagen alpha 1(I
86	81.0	920	2	A45748	collagen alpha 2(I
87	81.0	1018	1	CGHU2A	collagen alpha 2(I
88	81.0	1029	1	S21369	collagen alpha 2(I
89	81.0	1712	1	CGHU2B	collagen alpha 2(I
90	81.0	1744	2	S40991	collagen alpha 1(I
91	81.0	1747	2	A45974	collagen alpha 1(X
92	81.0	1758	2	T29350	hypothetical prote
93	81.0	1759	2	T29351	collagen alpha 1(X
94	81.0	1857	2	S31212	collagen alpha 1(X
95	81.0	1888	2	S78476	probable polyketid
96	81.0	2126	2	H70621	probable polyketid
97	81.0	4151	2	G70944	collagen 6, nonfib
98	81.0	118	2	B41207	hypothetical glyci
99	81.0	175	2	A70824	cyclin-dependent k
100	81.0	181	2	I54380	cyclin-dependent k
101	81.0	181	2	I68674	hypothetical prote
102	81.0	208	2	T15245	hypothetical prote

103	32	76.2	210	2	A37358	oncofetal protein	176	32	76.2	902	2	T26775	hypothetical prote
104	32	76.2	220	2	A55993	hypothetical prote	177	32	76.2	914	2	H70987	hypothetical glyci
105	32	76.2	224	2	I38940	zinc finger protei	178	32	76.2	957	2	D70835	hypothetical glyci
106	32	76.2	266	2	T22706	hypothetical prote	179	32	76.2	963	2	T19140	hypothetical prote
107	32	76.2	276	2	T19845	hypothetical prote	180	32	76.2	1007	2	E72489	hypothetical prote
108	32	76.2	283	2	T29837	hypothetical prote	181	32	76.2	1019	1	A32856	collagen alpha 1(V
109	32	76.2	283	2	T19141	hypothetical prote	182	32	76.2	1051	2	A35763	collagen alpha 2 c
110	32	76.2	283	2	T32921	hypothetical prote	183	32	76.2	1120	2	H88449	protein F5408.1 [i
111	32	76.2	286	2	B45632	marzoite surface	184	32	76.2	1142	2	JK0369	collagen alpha 1(X
112	32	76.2	291	2	T20942	hypothetical prote	185	32	76.2	1298	1	EDB875	immediate-early pr
113	32	76.2	294	2	T29838	hypothetical prote	186	32	76.2	1315	2	A56101	collagen alpha 1(X
114	32	76.2	294	2	T29839	hypothetical prote	187	32	76.2	1366	1	CGHU2S	collagen alpha 2(I
115	32	76.2	294	2	T21668	hypothetical prote	188	32	76.2	1373	1	A43291	collagen alpha 2(I
116	32	76.2	294	2	T22639	hypothetical prote	189	32	76.2	1400	2	T31555	hypothetical prote
117	32	76.2	295	2	A44994	collagen - nematod	190	32	76.2	1433	2	T30261	chitinase (BC 3.2.
118	32	76.2	299	2	AH3447	cytochrome-c oxida	191	32	76.2	1497	2	I49607	procollagen type V
119	32	76.2	299	2	T19584	hypothetical prote	192	32	76.2	1546	1	CGHU2E	collagen alpha 2(X
120	32	76.2	299	2	T22705	hypothetical prote	193	32	76.2	1572	2	T00027	brain-specific ang
121	32	76.2	299	2	T24833	hypothetical prote	194	32	76.2	1691	1	CGHU6B	collagen alpha 6(I
122	32	76.2	299	2	T25407	hypothetical prote	195	32	76.2	1733	1	B45344	probable nuclear a
123	32	76.2	300	2	T24482	hypothetical prote	196	32	76.2	1755	2	S61763	TyB protein - yeas
124	32	76.2	301	2	B31219	collagen 2 - Caeno	197	32	76.2	1755	2	S69845	TyB protein - yeas
125	32	76.2	302	2	A31921	collagen dpy-13 pr	198	32	76.2	1756	2	S45867	TyB protein - yeas
126	32	76.2	304	2	T22482	hypothetical prote	199	32	76.2	1774	2	B56101	collagen alpha 1(X
127	32	76.2	304	2	T26185	hypothetical prote	200	32	76.2	1783	2	T37258	probable voltage-d
128	32	76.2	304	2	T26184	hypothetical prote	201	32	76.2	1917	2	C88728	protein C48A7.1 [i
129	32	76.2	305	2	T30165	hypothetical prote	202	32	76.2	1958	2	B40505	hypothetical prote
130	32	76.2	307	2	T18846	hypothetical prote	203	32	76.2	2875	1	RRVUTW	genome polypeptid
131	32	76.2	310	2	T22641	hypothetical prote	204	32	76.2	3198	2	A43426	collagen alpha 2 f
132	32	76.2	312	2	T25048	hypothetical prote	205	31	73.8	58	2	AF1845	hypothetical prote
133	32	76.2	333	2	T20436	hypothetical prote	206	31	73.8	78	2	B44479	collagen alpha 1(X
134	32	76.2	341	2	T16296	hypothetical prote	207	31	73.8	108	2	A46222	hydrophobin Ccg-2
135	32	76.2	345	2	T29981	hypothetical prote	208	31	73.8	144	2	A83313	probable type II s
136	32	76.2	348	2	T29288	hypothetical prote	209	31	73.8	149	2	G83560	probable type II s
137	32	76.2	350	2	H87272	3-isopropylmalate	210	31	73.8	166	2	G97522	hypothetical prote
138	32	76.2	369	2	S33603	surfactant protein	211	31	73.8	166	2	AH2741	hypothetical prote
139	32	76.2	371	1	JN0450	conglutinin precur	212	31	73.8	171	2	A34493	collagen alpha 1(I
140	32	76.2	371	2	I45878	conglutinin - bovi	213	31	73.8	196	2	G86313	hypothetical prote
141	32	76.2	372	2	A52613	DNA polymerase III	214	31	73.8	196	2	G83757	ribonuclease H-rel
142	32	76.2	375	1	A45225	pulmonary surfacta	215	31	73.8	210	2	T20973	hypothetical prote
143	32	76.2	383	2	D75081	phosphate abc tran	216	31	73.8	237	2	A88640	protein C34H4.4 [i
144	32	76.2	403	2	C97395	DNA polymerase III	217	31	73.8	232	2	A55047	collagen alpha 1(V
145	32	76.2	427	2	A43453	A-kinase anchor pr	218	31	73.8	253	2	T08152	axonal inner dyn
146	32	76.2	438	2	S53787	collagen alpha cha	219	31	73.8	281	2	T32765	hypothetical prote
147	32	76.2	460	2	A31897	muscarinic acetyl	220	31	73.8	284	2	T28887	collagen dpy-10 -
148	32	76.2	460	2	I51837	muscarinic recepto	221	31	73.8	285	2	S60598	collagen alpha cha
149	32	76.2	460	2	S09508	muscarinic acetyl	222	31	73.8	289	2	T20177	hypothetical prote
150	32	76.2	460	2	A24325	muscarinic acetyl	223	31	73.8	291	2	T34494	hypothetical prote
151	32	76.2	460	2	A29514	muscarinic acetyl	224	31	73.8	295	2	F83201	conserved hypotet
152	32	76.2	460	2	T33110	hypothetical prote	225	31	73.8	295	2	T19220	hypothetical prote
153	32	76.2	461	2	F70571	hypothetical glyci	226	31	73.8	301	2	T23441	hypothetical prote
154	32	76.2	469	2	A24450	collagen alpha 2(V	227	31	73.8	302	2	T32872	hypothetical prote
155	32	76.2	482	2	S76475	hypothetical prote	228	31	73.8	304	2	T16107	hypothetical prote
156	32	76.2	487	2	E70983	hypothetical glyci	229	31	73.8	306	2	T21939	hypothetical prote
157	32	76.2	492	2	S28139	gas vesicle protei	230	31	73.8	309	2	T28708	hypothetical prote
158	32	76.2	536	2	S15183	gas-vesicle operon	231	31	73.8	312	1	Q0B807	US10 protein - hum
159	32	76.2	536	2	T08241	gas-vesicle operon	232	31	73.8	313	2	T26465	hypothetical prote
160	32	76.2	545	2	S28117	gas-vesicle operon	233	31	73.8	317	2	T29960	hypothetical prote
161	32	76.2	576	2	A70900	hypothetical glyci	234	31	73.8	320	1	A39724	homeotic protein H
162	32	76.2	580	2	E87532	alpha-amylase fami	235	31	73.8	324	2	T28032	hypothetical prote
163	32	76.2	584	2	G70804	hypothetical glyci	236	31	73.8	326	2	T16841	hypothetical prote
164	32	76.2	603	2	A70770	hypothetical glyci	237	31	73.8	330	2	T26004	hypothetical prote
165	32	76.2	615	2	B41873	dnaK-type molecula	238	31	73.8	346	2	AC3108	hypothetical prote
166	32	76.2	615	2	B97058	molecular chaperon	239	31	73.8	363	2	T16831	hypothetical prote
167	32	76.2	615	2	H70589	hypothetical glyci	240	31	73.8	371	2	A98179	ABC transporter, A
168	32	76.2	623	2	A45137	collagen alpha 4(I	241	31	73.8	375	2	H83290	conserved hypotet
169	32	76.2	677	2	S23296	collagen alpha 2(I	242	31	73.8	380	2	A48295	collagen 1 - marin
170	32	76.2	735	2	S54147	alpha adducin - ra	243	31	73.8	386	2	T27079	hypothetical prote
171	32	76.2	737	2	S18207	aducan alpha chai	244	31	73.8	388	2	T36402	probable monooxyge
172	32	76.2	801	2	F70824	hypothetical glyci	245	31	73.8	397	2	AH3494	DNA-directed DNA p
173	32	76.2	837	2	E70835	hypothetical glyci	246	31	73.8	402	1	QRBOT2	microtubule-associ
174	32	76.2	859	2	T35785	probable beta-gluc	247	31	73.8	418	2	T15142	hypothetical prote
175	32	76.2	889	2	B55123	coatomer complex b	248	31	73.8	428	2	T24769	hypothetical prote

249	31	73.8	437	2	T42653	hypothetical prote	322	30	71.4	314	2	T34269	hypothetical prote
250	31	73.8	448	1	OR80T1	microtubule-associ	323	30	71.4	323	2	A35909	fra-2 protein - ch
251	31	73.8	452	2	T30082	hypothetical prote	324	30	71.4	323	2	T27450	hypothetical prote
252	31	73.8	484	2	G87582	cobryic acid synth	325	30	71.4	330	2	S46657	collagen alpha 1(X
253	31	73.8	484	2	C87466	cobryic acid synth	326	30	71.4	337	2	B54497	surface membrane p
254	31	73.8	518	2	A55840	macrophage bacteri	327	30	71.4	339	2	T22607	hypothetical prote
255	31	73.8	532	2	T70580	hypothetical glyci	328	30	71.4	348	2	A34705	collagen - Caenorh
256	31	73.8	624	2	A55576	collagen alpha 2(X	329	30	71.4	356	2	S16907	collagen alpha 1(I
257	31	73.8	629	2	A30168	homeotic protein L	330	30	71.4	357	2	S35250	flagellar biosynth
258	31	73.8	635	2	A57131	collagen alpha 2(V	331	30	71.4	377	2	A47380	RING finger-contai
259	31	73.8	635	2	S01164	homeotic protein L	332	30	71.4	382	2	S20375	collagen alpha 3(V
260	31	73.8	667	2	A97376	ba-type ubiquinol	333	30	71.4	402	1	C8B02S	collagen alpha 2(I
261	31	73.8	667	2	AG2593	cytochrome O ubiq	334	30	71.4	404	2	C75027	dihydroorotase (py
262	31	73.8	674	2	S23297	collagen alpha 1(X	335	30	71.4	417	2	C71212	probable dihydroor
263	31	73.8	674	2	S13301	collagen alpha 1(X	336	30	71.4	427	2	F64064	tolB protein - Hae
264	31	73.8	705	2	T38452	probable transmemb	337	30	71.4	438	2	D90734	probable tail fibre
265	31	73.8	722	2	S57246	ventral nervous sy	338	30	71.4	439	2	D70954	hypothetical glyci
266	31	73.8	741	2	B82236	isocitrate dehydro	339	30	71.4	453	2	S18804	collagen alpha 4(I
267	31	73.8	757	2	F83516	hypothetical prote	340	30	71.4	464	2	S59513	collagen alpha 4(I
268	31	73.8	775	2	A61228	collagen alpha 2(I	341	30	71.4	464	2	T36256	collagen 11 Al pro
269	31	73.8	778	2	F70963	hypothetical glyci	342	30	71.4	471	2	A39024	probable membrane
270	31	73.8	827	2	T40394	conserved hypothet	343	30	71.4	480	2	JH0672	brain factor 1 pro
271	31	73.8	918	2	S23377	collagen alpha 2(V	344	30	71.4	488	2	A27353	collagen alpha 1(I
272	31	73.8	920	2	B34493	collagen alpha 1(I	345	30	71.4	498	2	C70720	hypothetical glyci
273	31	73.8	931	2	S13580	collagen alpha 1(I	346	30	71.4	504	2	E87628	hypothetical prote
274	31	73.8	1022	2	S04111	collagen alpha 2(V	347	30	71.4	505	2	S54066	hypothetical prote
275	31	73.8	1106	2	JQ0405	collagen alpha 2(V	348	30	71.4	509	2	H87389	conserved hypothet
276	31	73.8	1274	2	JN0015	hypothetical 119.5	349	30	71.4	515	2	H70663	hypothetical glyci
277	31	73.8	1275	2	JU0092	trp protein - frui	350	30	71.4	516	2	C44479	collagen alpha 1(X
278	31	73.8	1340	2	A39808	trp protein - frui	351	30	71.4	518	2	T23745	hypothetical prote
279	31	73.8	1761	2	T13990	proteoglycan core	352	30	71.4	523	2	S03572	DNA-directed RNA p
280	31	73.8	1838	1	CGHU1V	collagen type IV a	353	30	71.4	523	2	B84416	DNA-directed RNA p
281	31	73.8	1843	2	S18803	collagen alpha 1(V	354	30	71.4	537	2	C90389	conserved hypothet
282	31	73.8	2327	2	T42630	collagen alpha 1(V	355	30	71.4	544	2	T36645	probable large int
283	30	71.4	52	2	S69113	aggrecan - bovine	356	30	71.4	547	2	A36046	collagen alpha cha
284	30	71.4	54	2	A26463	collagen alpha 3(I	357	30	71.4	554	2	AE0584	asparagine synthet
285	30	71.4	81	2	A49736	relaxin - spiny do	358	30	71.4	564	1	HMIVF2	hemagglutinin prec
286	30	71.4	101	1	HMIVH4	collagen alpha 3(I	359	30	71.4	564	1	HMIVF3	hemagglutinin prec
287	30	71.4	128	2	S47036	hemagglutinin prec	360	30	71.4	564	1	HMIVF4	hemagglutinin prec
288	30	71.4	142	2	S51582	structural protein	361	30	71.4	564	1	HMIVF5	hemagglutinin prec
289	30	71.4	177	2	S37749	galactoside 2-alpha	362	30	71.4	564	1	HMIVF6	hemagglutinin prec
290	30	71.4	198	2	S21388	collagen alpha 2(X	363	30	71.4	564	1	HMIVF7	hemagglutinin prec
291	30	71.4	205	1	QWHDUX	type I keratin 48k	364	30	71.4	564	1	HMIVF8	hemagglutinin prec
292	30	71.4	212	2	A72648	lymphotoxin alpha	365	30	71.4	564	1	HMIVF1	hemagglutinin prec
293	30	71.4	214	2	B34503	hypothetical prote	366	30	71.4	568	1	A46339	hemagglutinin prec
294	30	71.4	220	2	B49736	small nuclear ribo	367	30	71.4	581	2	T08846	asparagine synthas
295	30	71.4	228	2	A44982	collagen alpha 3(I	368	30	71.4	583	1	AJFWM2	ABC transporter, A
296	30	71.4	231	2	I53659	Sm-B protein - mou	369	30	71.4	583	2	H75561	asparagine synthas
297	30	71.4	237	2	T35351	hypothetical prote	370	30	71.4	584	2	T12989	asparagine synthas
298	30	71.4	240	2	A13182	conserved hypothet	371	30	71.4	586	2	S52387	asparagine synthas
299	30	71.4	240	2	S09377	small nuclear ribo	372	30	71.4	586	2	S69182	2,3-dihydroxybenzo
300	30	71.4	245	1	C1HQQA	complement subcomp	373	30	71.4	587	2	AD3519	malate oxidoreduct
301	30	71.4	245	2	F70787	hypothetical glyci	374	30	71.4	588	2	A82332	asparagine synthas
302	30	71.4	259	2	A86507	biotin protein lig	375	30	71.4	590	2	S25165	asparagine synthas
303	30	71.4	259	2	A72116	biotin apo-protein	376	30	71.4	591	2	I48141	acroganin - guine
304	30	71.4	264	2	E83284	probable enoyl-CoA	377	30	71.4	591	2	B70523	hypothetical glyci
305	30	71.4	275	2	D83536	hypothetical prote	378	30	71.4	602	2	F84432	probable C2H2-type
306	30	71.4	281	2	C88638	protein F58f6.1 li	379	30	71.4	614	2	T33149	hypothetical prote
307	30	71.4	282	2	J50168	collagen col-8 - C	380	30	71.4	615	2	A05269	collagen alpha 1(I
308	30	71.4	282	2	T16036	cuticle collagen c	381	30	71.4	627	2	G96537	hypothetical prote
309	30	71.4	283	2	T19731	hypothetical prote	382	30	71.4	637	1	KXMSC2	proprotein convert
310	30	71.4	283	2	T19732	hypothetical prote	383	30	71.4	637	1	KXRTC2	proprotein convert
311	30	71.4	283	2	T29980	hypothetical prote	384	30	71.4	638	2	S23244	proprotein convert
312	30	71.4	284	2	T29528	hypothetical prote	385	30	71.4	638	2	S23118	mannosyl-oligosacc
313	30	71.4	285	2	T29982	hypothetical prote	386	30	71.4	639	2	S78554	neurexin II-beta p
314	30	71.4	289	2	T26812	hypothetical prote	387	30	71.4	659	2	D40228	collagen alpha 1(I
315	30	71.4	296	2	T24827	hypothetical prote	388	30	71.4	662	2	C8B06C	collagen alpha 1(I
316	30	71.4	301	2	A53570	collectin-43 - bov	389	30	71.4	673	1	CNHUB	chromogranin B pre
317	30	71.4	302	2	T21257	hypothetical prote	390	30	71.4	677	1	JCT303	pectate lyase (EC
318	30	71.4	306	2	T51705	methylenetetrahydr	391	30	71.4	677	2	CGHUID	collagen alpha 1(X
319	30	71.4	307	2	T34973	5,10-methylenetet	392	30	71.4	680	1	CGHUID	collagen alpha 1(X
320	30	71.4	308	2	T37286	collagen 40 - Caen	393	30	71.4	684	2	A53019	collagen alpha 1(X
321	30	71.4	311	2	T15268	hypothetical prote	394	30	71.4	688	2	A53330	collagen alpha 2(I

395 30 71.4 708 2 T00205 hypothetical prote
396 30 71.4 714 2 T22454 scavenger receptor
397 30 71.4 742 2 J7595 collagen alpha 1(V
398 30 71.4 743 1 S23779 collagen alpha 1(V
399 30 71.4 744 1 A34246 collagen alpha 1(V
400 30 71.4 744 1 S23298 collagen alpha 1(V
401 30 71.4 744 2 S15435 chitinase - Strept
402 30 71.4 765 2 T35719 LDL receptor relat
403 30 71.4 770 2 T00204 LDL receptor-relat
404 30 71.4 770 2 T00203 hypothetical glyci
405 30 71.4 783 2 E70824 cytokine receptor
406 30 71.4 897 1 A39255 hypothetical prote
407 30 71.4 904 2 T22457 probable rne prote
408 30 71.4 953 2 B70681 collagen alpha 3(V
409 30 71.4 959 2 S32605 hypothetical prote
410 30 71.4 962 2 T22459 transposase - Bseu
411 30 71.4 988 2 A24341 transposase - Bseu
412 30 71.4 988 2 JQ1477 transposase (EC 6.
413 30 71.4 988 4 S58114 collagen alpha 1(V
414 30 71.4 1025 2 S34839 collagen alpha 1(V
415 30 71.4 1028 1 CGHUIA scavenger receptor
416 30 71.4 1036 2 T17405 collagen, cornea-s
417 30 71.4 1146 2 A38587 stripe a/b protein
418 30 71.4 1180 2 S69205 hypothetical prote
419 30 71.4 1230 2 T22458 kinesi-related pr
420 30 71.4 1265 2 T03792 kinesin heavy chai
421 30 71.4 1265 2 T07397 probable immediate
422 30 71.4 1300 2 T03166 probable surface 1
423 30 71.4 1331 2 A72647 collagen alpha 1(X
424 30 71.4 1388 2 A53317 collagen alpha 2(V
425 30 71.4 1496 1 CGHU2V collagen alpha 1(X
426 30 71.4 1532 2 A61262 hypothetical prote
427 30 71.4 1662 2 T01893 neurexin II-alpha
428 30 71.4 1715 2 C40228 GII protein - mous
429 30 71.4 1840 2 T30250 mannose 6-phosphat
430 30 71.4 2499 1 A30788 erythronolide synt
431 30 71.4 3172 2 S22012 collagen alpha 3(V
432 30 71.4 3176 2 CGHU3A 6-deoxyerythronol
433 30 71.4 3178 2 S13595 leukocyte chemoatt
434 29 69.0 16 2 J70609 collagen alpha 1(I
435 29 69.0 43 2 S13581 hypothetical prote
436 29 69.0 86 2 G70611 oncofetal-laminin
437 29 69.0 92 2 A38947 collagen alpha 1(I
438 29 69.0 93 2 I45876 hypothetical prote
439 29 69.0 94 2 B86195 hypothetical prote
440 29 69.0 96 2 C82518 peptidylprolyl iso
441 29 69.0 107 2 A61431 peptidylprolyl iso
442 29 69.0 108 2 A42657 FK506-binding prot
443 29 69.0 108 2 J34578 FK 506-binding pro
444 29 69.0 108 2 J50764 FK506-binding prot
445 29 69.0 108 2 JH0528 brain and muscle A
446 29 69.0 113 2 PC4289 anther specific pr
447 29 69.0 128 2 T03603 7,8-dihydro-8-oxog
448 29 69.0 129 1 MVECMT 7,8-dihydro-8-oxog
449 29 69.0 131 2 A70519 7,8-dihydro-8-oxog
450 29 69.0 132 2 G85492 7,8-dihydro-8-oxog
451 29 69.0 132 2 G30641 7,8-dihydro-8-oxog
452 29 69.0 133 2 E70795 hypothetical prote
453 29 69.0 134 2 D83127 hypothetical prote
454 29 69.0 137 2 H70811 hypothetical glyci
455 29 69.0 139 2 T34633 hypothetical prote
456 29 69.0 157 2 T10694 hypothetical prote
457 29 69.0 161 2 S12246 anther-specific pr
458 29 69.0 164 2 F84286 hypothetical prote
459 29 69.0 165 2 A52739 acetyltransferase
460 29 69.0 165 2 C97520 hypothetical prote
461 29 69.0 183 2 E90141 oxidoreductase, so
462 29 69.0 197 1 JH0309 tumor necrosis fac
463 29 69.0 201 2 T07011 proteinase inhibit
464 29 69.0 205 2 T33781 hypothetical prote
465 29 69.0 205 2 T34724 probable membrane
466 29 69.0 208 2 C34503 small nuclear ribo
467 29 69.0 209 2 F87394 isochorismatase fa

468 29 69.0 210 2 T03803 tumor-related prot
469 29 69.0 210 2 B44984 collagen - nematod
470 29 69.0 212 2 J75511 fibroblast growth
471 29 69.0 219 2 T14782 hypothetical prote
472 29 69.0 220 2 AD2890 conserved hypothet
473 29 69.0 220 2 E98293 hypothetical prote
474 29 69.0 221 2 A84596 hypothetical prote
475 29 69.0 222 2 A88102 protein W09G10.1 |
476 29 69.0 227 2 S50067 homeotic protein H
477 29 69.0 229 2 E83135 probable dehydroge
478 29 69.0 237 1 C69040 probable 3-isoprop
479 29 69.0 239 1 LNMSWA mannose-binding le
480 29 69.0 240 2 A33270 snRNP protein N -
481 29 69.0 240 2 S20068 small nuclear ribo
482 29 69.0 240 2 A34503 crystallin beta B1
483 29 69.0 241 2 S53522 hypothetical prote
484 29 69.0 242 2 C83045 gelatin-binding 28
485 29 69.0 244 2 JC4708 complement subcomp
486 29 69.0 245 1 C1HUQC complement subcomp
487 29 69.0 245 2 S19018 transforming prote
488 29 69.0 248 1 TVMVR5 pulmonary surfacta
489 29 69.0 248 1 LNRTPS pulmonary surfacta
490 29 69.0 248 2 A48853 hypothetical prote
491 29 69.0 249 2 E82638 hypothetical prote
492 29 69.0 250 2 A70678 complement subcomp
493 29 69.0 253 1 C1HUQB hypothetical glyci
494 29 69.0 256 2 A70514 ABC transporter AT
495 29 69.0 258 2 T35444 hydrolase, CbbY/Cb
496 29 69.0 259 2 H75481 collagen, alpha 2(I
497 29 69.0 261 2 A34476 hypothetical prote
498 29 69.0 263 2 C72100 hypothetical prote
499 29 69.0 263 2 C86524 hypothetical prote
500 29 69.0 265 2 S39874 hypothetical prote

ALIGNMENTS

RESULT 1

B40983

collagen alpha 1(XIII) chain precursor - human

N;Alternate names: low molecular weight collagen alpha chain; procollagen alpha 1(XIII)
N;Contains: procollagen alpha 1(XIII) chain, splice form A; procollagen alpha 1(XIII) ch

m D

C;Species: Homo sapiens (man)

C;Date: 13-May-1992 #sequence revision 17-Nov-1995 #text change 09-Jul-2004

C;Accession: B40983; A40983; A38298; B38298; C38298; A26412; B26412; I59133; I79536; I79

R;Titka, L.; Elomaa, O.; Pihlajaniemi, T.; Tryggvason, K.

J. Biol. Chem. 266, 17713-17719, 1991

A;Title: Human alpha1(XIII) collagen gene. Multiple forms of the gene transcripts are ge

A;Reference number: A40983; MUID:91373404; PMID:1894651

A;Accession: B40983

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-41,'X',48-53,'X',60-374,'L',376-446,'Q',448-633 <TIK>

A;Cross-references: UNIPROT:Q13992; UNIPROT:Q13995; UNIPROT:Q13994; UNIPROT:Q13993; UNIP

AA52754.1; PID:G180756

A;Accession: A40983

A;Molecule type: DNA

A;Residues: 381-401,416-466 <TI2>

A;Cross-references: UNIPARC:UPI000017A183; GB:M68995; GB:M68997

R;Pihlajaniemi, T.; Tamminen, M.

J. Biol. Chem. 265, 16922-16928, 1990

A;Title: The alpha1 chain of type XIII collagen consists of three collagenous and four m

A;Reference number: A38298; MUID:91009112; PMID:1698771

A;Accession: A38298

A;Molecule type: mRNA

A;Residues: 1-50,80-98,'P',100-155,178-310,'A',312-382,'L',384-385,'L',387-391,'L',393-3

-633 <PIH>

A;Cross-references: UNIPARC:UPI000017A184; GB:J05580; GB:M59217; NID:G178319

A;Note: splice form E-3; the authors translated the codon GCG for residue 353 as Gly

A;Accession: B38298

A;Molecule type: mRNA

A;Accession: B29301
A:Molecule type: mRNA
A;Residues: 1481-1707 <KUR>
A;Cross-references: UNIPARC:UPI000016CC4C; EMBL:M15933; NID:g192284; PIDN:AAA37341.1; PI
R;Schwarz-Magdolen, U.; Oberbaeumer, I.; Kuehn, K.
FEBS Lett. 208, 203-207, 1986
A;Title: cDNA and protein sequence of the NC1 domain of the alpha-2-chain of collagen IV
A;Reference number: A24628; MUID:87054581; PMID:3780963
A;Accession: A24628
A:Molecule type: mRNA
A;Residues: 1480-1572, 'L', 1574-1622, 'H', 1624-1707 <SCH>
A;Cross-references: UNIPARC:UPI00001773BA; EMBL:X04647
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988
A;Title: Head-to-head arrangement of murine type IV collagen genes.
A;Reference number: A92702; MUID:89066738; PMID:3198626
A;Accession: B32003
A:Molecule type: DNA
A;Residues: 1-33 <KA2>
A;Cross-references: UNIPARC:UPI000016CCAF; EMBL:J04448; NID:g192666; PIDN:AAA37438.1; PI
R;Burbello, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
A;Title: Alpha1(IV) and Alpha2(IV) collagen genes are regulated by a bidirectional prom
A;Reference number: A94220; MUID:89071759; PMID:3200851
A;Accession: B31766
A:Molecule type: DNA
A;Residues: 1-60 <BUR>
A;Cross-references: UNIPARC:UPI000016CCB3; EMBL:M23333
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A;Title: Subunit structure and assembly of the globular
A;Reference number: S17801; MUID:84132058; PMID:6698021
A;Accession: S19086
A:Molecule type: protein
A;Residues: 1475-1481, 'X', 1483-1487 <WEB>
A;Cross-references: UNIPARC:UPI00001773BB
C;Genetics:
A;Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3
C;Superfamily: collagen alpha 1(IV) chain
A;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;29-171/Domain: 7S #status predicted <7SD>
F;58-1480/Domain: collagenous #status predicted <COL>
F;141-143/Region: cell attachment (R-G-D) motif
F;360-362/Region: cell attachment (R-G-D) motif
F;779-781/Region: cell attachment (R-G-D) motif
F;884-886/Region: cell attachment (R-G-D) motif
F;965-967/Region: cell attachment (R-G-D) motif
F;1223-1225/Region: cell attachment (R-G-D) motif
F;1447-1449/Region: cell attachment (R-G-D) motif
F;1481-1707/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1481-1589/Domain: repeat NC1 #status predicted <NC11>
F;1590-1707/Domain: repeat NC1 #status predicted <NC12>
F;42,47,51,53,481,483/Disulfide bonds: interchain #status predicted
F;138,1370/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;656-676,544-1550,1653-1660/Disulfide bonds: #status predicted
F;985,988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Bind
F;985,988,997,1003,1028,1031,1067,1101,1113,1119,1143,1170,1200,1242,1305,1368,1391/Modi
F;1009,1012,1018,1021,1024,1037,1040,1043,1046,1052,1058,1070,1098,1110,1128,1140,1149,1
77,1383,1386,1401,1408,1420,1423,1429,1444,1465,1468,1471,1477/Modified site: hydroxypro

Query Match 92.9%; Score 39; DB 2; Length 1707;
Best Local Similarity 87.5%; Pred. No. 51;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 1135 GAQGSPL 1142
|||:|||||
RESULT 3
149558

collagen alpha 1(I) precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C;Accession: 149558
R;Monson, J.M.; McCarthy, B.J.
DNA 1, 59-69, 1981
A;Title: Identification of a Balb/c mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: 149558; MUID:83157109; PMID:6219867
A;Accession: 149558
A;Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A;Residues: 1-198 <RES>
A;Cross-references: UNIPARC:UPI000016CC45; GB:K03036; NID:g192258; PIDN:AAA37332.1; PID:
C;Genetics:
A;Gene: COL1A1
A;Introns: 36/3; 54/3; 72/3; 108/3; 126/3; 144/3; 162/3
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
Query Match 90.5%; Score 38; DB 2; Length 198;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPG 7
Db 148 GAEGSPG 154
|||||
RESULT 4
CGBOLIS
collagen alpha 1(I) chain - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1994 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A91193; A91229; A91387; A91211; A91201; A91200; A43048; A02853
R;Rauterberg, J.; Timpl, R.; Furthmayr, H.
Eur. J. Biochem. 27, 231-237, 1972
A;Title: Structural characterization of N-terminal antigenic determinants in calf and hu
A;Reference number: A91193; MUID:72255334; PMID:4115172
A;Accession: A91193
A:Molecule type: protein
A;Residues: 1-19 <RAU>
A;Cross-references: UNIPROT:P02453; UNIPARC:UPI00000173B51
A;Experimental source: skin
A;Note: the epsilon carbon of Lys-9, by homology with the rat alpha 1(I) chain, is conve
R;Fietzek, P.P.; Kuehn, K.
Eur. J. Biochem. 52, 77-82, 1975
A;Title: The covalent structure of collagen: amino-acid sequence of the cyanogen-bromide
A;Reference number: A91229; MUID:76022320; PMID:1164916
A;Accession: A91229
A:Molecule type: protein
A;Residues: 20-145 <FIE>
A;Cross-references: UNIPARC:UPI00000173B52
A;Experimental source: skin
A;Note: Lys-103 is hydroxylated and binds glucosylgalactose
R;Fietzek, P.P.; Wendt, P.; Kell, I.; Kuehn, K.
FEBS Lett. 26, 74-76, 1972
A;Title: The covalent structure of collagen: amino acid sequence of alpha1-CB3 from calf
A;Reference number: A91387; MUID:73049499; PMID:4673951
A;Accession: A91387
A:Molecule type: protein
A;Residues: 146-294 <F12>
A;Cross-references: UNIPARC:UPI00000173B53
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Hopper, K.E.; Kuehn, K.
Eur. J. Biochem. 38, 398-400, 1973
A;Title: The covalent structure of collagen. 2. The amino-acid sequence of alpha1-CB7 fr
A;Reference number: A91211; MUID:74086118; PMID:4359390
A;Accession: A91211
A:Molecule type: protein
A;Residues: 295-562 <F13>
A;Cross-references: UNIPARC:UPI00000173B54
A;Experimental source: skin
R;Wendt, P.; Mark, K.V.D.; Rexrodt, F.; Kuehn, K.
Eur. J. Biochem. 30, 169-183, 1972

A;Title: The covalent structure of collagen. The amino-acid sequence of the 112 residues
A;Reference number: A91201; MUID:73042276; PMID:4343808
A;Accession: A91201
A;Molecule type: protein
A;Residues: 563-675 <WEN>
A;Cross-references: UNIPARC:UPI0000173B55
A;Experimental source: skin
R;Fietzek, P.P.; Rexrodt, F.W.; Wendt, P.; Stark, M.; Kuehn, K.
Eur. J. Biochem. 30, 163-168, 1972
A;Title: The covalent structure of collagen. Amino acid sequence of peptide alpha1-CB6-C
A;Reference number: A91200; MUID:73042275; PMID:4343807
A;Accession: A91200
A;Molecule type: protein
A;Residues: 676-758 <FIA>
A;Cross-references: UNIPARC:UPI0000173B56
A;Experimental source: skin
R;Rautenberg, J.; Fietzek, P.; Rexrodt, F.; Becker, U.; Stark, M.; Kuehn, K.
FEBS Lett. 21, 75-79, 1972
A;Title: The amino acid sequence of the carboxyterminal nonhelical cross link region of
A;Reference number: A43048
A;Accession: A43048
A;Molecule type: protein
A;Residues: 759-779 <RA2>
A;Cross-references: UNIPARC:UPI0000173B57
A;Experimental source: skin
C;Comment: Lysines at positions 115, 124, 274, 346, 424, 496, 658, and 670 may be hydrox
C;Comment: Prolines in the third position of the tripeptide repeating unit (G-X-Y) are h
C;Comment: The order of the eight CNBR peptides in the alpha 1(I) chain of bovine skin c
9, 149, 268, and 217 residues.
C;Comment: The complete chain contains 1052 residues.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; proglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 90.5%; Score 38; DB 1; Length 779;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 581 GAEGSPG 587

RESULT 5
S21626 collagen alpha 1(I) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I
A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>
A;Cross-references: UNIPROT:P11087; UNIPARC:UPI0000027558; EMBL:U08020; NID:G470673; PID
R;Metaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A;Reference number: S16176; MUID:91274355; PMID:2054384
A;Accession: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: UNIPARC:UPI0000000691; EMBL:X57981; NID:G50484; PIDN:CAA41046.1; PID
R;French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403; PMID:3841523
A;Accession: A23982
A;Molecule type: mRNA

A;Residues: 518-1128 <FRE>
A;Cross-references: UNIPARC:UPI000016CC46; GB:M14423; NID:G192261; PIDN:AAA37333.1; PID
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A;Reference number: I49559; MUID:83141374; PMID:6298597
A;Accession: I49559
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 735-1130 <RES>
A;Cross-references: UNIPARC:UPI000016CC47; GB:M17491; NID:G192263; PIDN:AAA37334.1; PID
R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.
Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984
A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads
A;Reference number: I49557; MUID:84170331; PMID:6324198
A;Accession: I49557
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-25 <RE2>
A;Cross-references: UNIPARC:UPI00000E5B79; GB:K01688; NID:G192246; PIDN:AAA37330.1; PID
R;Fenton, S.P.; Laman, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.
Biochim. Biophys. Acta 1216, 469-474, 1993
A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.
A;Reference number: S39789; MUID:94092741; PMID:8268229
A;Accession: S39789
A;Molecule type: DNA
A;Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1
A;Cross-references: UNIPARC:UPI00000E6565; UNIPARC:UPI000017738A
R;Rhodes, K.; Rippe, R.A.; Umezawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.
Mol. Cell. Biol. 14, 5950-5960, 1994
A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indire
A;Reference number: I48300; MUID:94344105; PMID:8065328
A;Accession: I48300
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>
A;Cross-references: UNIPARC:UPI000017738B; EMBL:X54876; NID:G50486; PIDN:CAA38657.1; PI
C;Genetics: COL1A1
A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology,
C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F;30-89/Domain: von Willebrand factor type C repeat homology <WVC>
F;152-1453/Product: collagen alpha 1(I) chain #status predicted <WAT>
F;1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 90.5%; Score 38; DB 2; Length 1453;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 1008 GAEGSPG 1014

RESULT 6
CGHUIS collagen alpha 1(I) chain precursor - human
N;Alternate names: procollagen alpha 1(I) chain
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1981 #sequence revision 04-Oct-1996 #text change 31-Dec-2004
C;Accession: I60114; S01143; A93335; I55254; A39943; I5537; A35233; S09400; B90567; S1
5269; A29439; I53466; A02852; I37247
R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.
Gene 67, 105-115, 1988
A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five
A;Reference number: I60114; MUID:88329734; PMID:2843432
A;Accession: I60114
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-369, 'L', 371-589 <DAL>

A;Residues: 710-720,'E',722-737,'E',739-745 <WAL>
A;Cross-references: UNIPARC:UPI0000173B4B
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Motte
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the C
A;Reference number: 154365; MUID:95187161; PMID:7881420
A;Accession: I54365
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 746-766,'S',768-781 <FOR>
A;Cross-references: UNIPARC:UPI000016AGFA; GB:L47667; NID:gl009093; PIDN:AAB59576.1; PID
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A;Reference number: A47426; MUID:93352646; PMID:8349697
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1336,1339-1387,'R',1389-1464 <CHE>
A;Cross-references: UNIPARC:UPI000073A2A; GB:S64596; NID:q407589; PIDN:AAB27856.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCBI:P:136445)
A;Note: does not represent an experimentally determined sequence but three different mu
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Cross-references: UNIPARC:UPI0000173B4C
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276,'H',1278-1464 <CH5>
A;Cross-references: UNIPARC:UPI0000173B4D
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336,1339-1464 <CH6>
A;Cross-references: UNIPARC:UPI0000173B4E
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387,'R',1389-1464 <CH7>
A;Cross-references: UNIPARC:UPI0000173B4F
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andraessen, P.; Charbonneau, H.; Ni
J. Biol. Chem. 263, 14605-14607, 1988

Query Match 90.5%; Score 38; DB 1; Length 1464;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GARGSPG 7
DB 1019 GARGSPG 1025

RESULT 7
S31521
collagen COLF1 - freshwater sponge (Ephydatia muelleri)
C;Species: Ephydatia muelleri
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S31521
R;Exposito, J.Y.; van der Rest, M.; Garrone, R.
submitted to the EMBL Data Library, December 1992
A;Description: The complete intron/exon structure of E. muelleri collagen gene suggests
A;Reference number: S31521
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-812 <EXP>
A;Cross-references: UNIPROT:Q06452; UNIPARC:UPI000007D7E2; EMBL:X69818; NID:g429036; PID

Query Match 88.1%; Score 37; DB 2; Length 812;
Best Local Similarity 87.5%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GARGSPGL 8
DB 84 GARGSPGL 91

RESULT 8

S28774
collagen alpha chain - tube worm (Riftia pachyptila) (fragment)
C;Species: Riftia pachyptila
C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004
C;Accession: S28774; S22915; S17581
R;Mann, K.; Gall, F.; Timpl, R.
Eur. J. Biochem. 210, 839-847, 1992
A;Title: Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen fra
A;Reference number: S28774; MUID:93130909; PMID:1483468
A;Accession: S28774
A;Molecule type: protein
A;Residues: 1-95,'X',97-107,'X',109-191,'X',193-260,'X',262-278,'X',280-572,'X',574-611,
A;Cross-references: UNIPARC:UPI000017A179
R;Mann, K.; Gall, F.; Timpl, R.
submitted to the Protein Sequence Database, July 1992
A;Description: Amino acid sequence and cell adhesion activity of a fibril-forming colle
A;Reference number: S22915
A;Accession: S22915
A;Molecule type: protein
A;Residues: 1-95,'X',97-107,'X',109-191,'X',193-260,'X',262-278,'X',280-572,'X',574-611,
A;Cross-references: UNIPARC:UPI000017A179
A;Note: 903-proline modified to 4-hydroxyproline was also found
R;Gall, F.; Wiedemann, H.; Mann, K.; Kuehn, K.; Timpl, R.; Engel, J.
J. Mol. Biol. 221, 209-223, 1991
A;Title: Molecular characterization of cuticle and interstitial collagens from worms col
A;Reference number: S17581; MUID:92015209; PMID:1920405
A;Accession: S17581
A;Molecule type: protein
A;Residues: 8-45;525-545,'X',547-566,'X',568-572,'X',574-611,'X',613-618,'X',811-882 <G>
A;Cross-references: UNIPARC:UPI000017A17A; UNIPARC:UPI000017A17B; UNIPARC:UPI000017A17C
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
h 3- and 4-hydroxylated in the X-position. Lysines are 5-hydroxylated and subsequently
C;Complex: homotrimer
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; homotrimer; i
F;1-12/Domain: amino-terminal telopeptide (fragment) <NTE>
F;13-1023/Domain: collagenous #status experimental <COL>
F;1024-1027/Domain: carboxyl-terminal telopeptide (fragment) <CTE>
F;21-24,123,243,273,276,285,291,303,348,381,621,645/Modified site: 4-hydroxyproline (Pr
F;27-39,94,122,90,93,128,150,162,165,174,177,180,207,216,219,228,237,249,255,306,312,321,
F;711,714,717,723,744,759,774,783,792,816,843,849,855,861,867,888,894,915,945,954,963,96
F;53,161,165,416,551,647,773,815,1010,1013,1016,1019/Modified site: 3-hydroxyproline (P
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: 5-hydroxylysine (Ly
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: carboxylate (Lys)
F;183,342,546,567,939/Modified site: 5-hydroxylysine (Lys) #status experimental
F;351,933/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 88.1%; Score 37; DB 2; Length 1027;
Best Local Similarity 87.5%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GARGSPGL 8
DB 739 GARGSPGL 746

RESULT 9

S23809
collagen alpha 2(I) chain homolog - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: S23809
R;Exposito, J.Y.; d'Alessio, M.; Solursh, M.; Ramirez, F.
J. Biol. Chem. 267, 15559-15562, 1992
A;Title: Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) col
A;Reference number: S23809; MUID:92348411; PMID:1639795

A:Accession: S23809
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1414 <EXP>
A:Cross-references: UNIPROT:Q26634; UNIPARC:UPI000007BD61; EMBL:M92040; NID:gi61435; PID
C:Superfamily: collagen alpha 2(II) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:1207-1414/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 88.1%; Score 37; DB 1; Length 1414;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 324 GARGSPGL 331

RESULT 10
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence revision 19-Apr-1996 #text_change 09-Jul-2004
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Accession: S59856
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
A:Cross-references: UNIPROT:P08121; UNIPARC:UPI0000177386; EMBL:X52046
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOM>
A:Cross-references: UNIPARC:UPI00000295D6; EMBL:X52046; NID:G575321; PIDN:CAA36279.1; PI
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: UNIPARC:UPI000016CCAB; EMBL:X57983; NID:G50476; PIDN:CAA41048.1; PID
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3; 29
58/3; 673/3; 706/3; 742/3; 760/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 976/3
C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-154/Domain: propeptide #status predicted <PRO>
F:32-92/Domain: von Willebrand factor type C repeat homology <WVC>
F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 88.1%; Score 37; DB 2; Length 1464;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 716 GARGSPGL 723

RESULT 11
CGHUIB
collagen alpha 4(IV) chain precursor - human
N:Alternate names: procollagen alpha 4(IV) chain
C:Species: Homo sapiens (man)

C:Date: 06-Feb-1995 #sequence revision 03-Oct-1995 #text_change 09-Jul-2004
C:Accession: A55360; S36854; S28777
R:Leinonen, A.; Marlyama, M.; Mochizuki, T.; Tryggvason, K.; Reeders, S.T.
J. Biol. Chem. 269, 26172-26177, 1994
A:Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. Comp
A:Reference number: A55360; MUID:95014445; PMID:7523402
A:Accession: A55360
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-1690 <LEI>
A:Cross-references: UNIPROT:PS3420; UNIPARC:UPI0000000769; GB:X81053; NID:G574805; PIDN:
R:Suginoto, M.; Ohashi, T.; Yoshioka, H.; Mateuo, N.; Ninomiya, Y.
FEBS Lett. 330, 122-128, 1993
A:Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen cha
A:Reference number: S36854; MUID:93374047; PMID:8365481
A:Accession: S36854
A:Molecule type: DNA; mRNA
A:Residues: 1219-1658, 'FE', 1661-1690 <SUG>
A:Cross-references: UNIPARC:UPI000016A432; DDBJ:D17391; NID:G440365; PIDN:BAA04214.1; PI
R:Experimental source: whole eye
R:Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A:Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A:Reference number: S28777; MUID:93054733; PMID:1429714
A:Accession: S28777
A:Molecule type: DNA
A:Residues: 1407-1424, 'G', 1426-1430, 'A', 1432-1439, 'L', 1441-1507 <KAN>
A:Cross-references: UNIPARC:UPI0000173BFC; GB:L01475; GB:L01476
A:Note: the codons given for 1438-Asp (GAG) and 1443-Gly (GCA) are inconsistent with the
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (e
d and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A4
A:Cross-references: GDB:132673; OMIM:120131
A:Map position: 2q35-2q37
A:Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-l
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney g
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxyl
F:1-38/Domain: signal sequence #status predicted <SIG>
F:39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>
F:39-61/Domain: amino-terminal nonhelical, NH1 <NHI>
F:62-1466/Region: interrupted helical
F:94-96/Region: cell attachment (R-G-D) motif
F:145-147/Region: cell attachment (R-G-D) motif
F:189-191/Region: cell attachment (R-G-D) motif
F:310-312/Region: cell attachment (R-G-D) motif
F:724-726/Region: cell attachment (R-G-D) motif
F:785-787/Region: cell attachment (R-G-D) motif
F:989-991/Region: cell attachment (R-G-D) motif
F:1212-1214/Region: cell attachment (R-G-D) motif
F:1467-1690/Domain: carboxyl-terminal nonhelical, NCL <NCL>
F:1471-1569/Domain: collagen IV carboxyl-terminal repeat <CTL>
F:1579-1686/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:47, 52, 55, 57, 266, 400, 460, 492, 494, 668, 790, 828, 1095, 1131, 1294, 1317, 1375, 1407/Disulfide bo
F:142, 669/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1480-1566, 1513-1569/Disulfide bonds: (or 1480-1569, 1513-1566) #status predicted
F:1525-1531, 1634-1641/Disulfide bonds: #status predicted
F:1588-1683, 1622-1686/Disulfide bonds: (or 1588-1686, 1622-1683) #status predicted

Query Match 88.1%; Score 37; DB 1; Length 1690;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 533 GAEGPPGL 540

```
RESULT 12
A1537
DNA-binding protein pAT 133 - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C:Accession: A41537; 138101; S29992
R:Mueller, H.J.; Skerka, C.; Bialonski, A.; Zipfel, P.F.
Proc. Natl. Acad. Sci. U.S.A. 88, 10079-10083, 1991
A:Title: Clone pAT 133 identifies a gene that encodes another human member of a class of
A:Reference number: A41537; MUID:92052214; PMID:1658795
A:Accession: A41537
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-486 <NUE>
A:Cross-references: UNIPROT:Q05215; UNIPARC:UPI0000161B28; GB:X60104; NID:G433902; PIDN:
R:Holst, C.; Skerka, C.; Lichter, P.; Bialonski, A.; Zipfel, P.F.
Hum. Mol. Genet. 2, 367-372, 1993
A:Title: Genomic organization, chromosomal localization and promoter function of the hum
A:Reference number: I38101; MUID:93278383; PMID:8504297
A:Accession: I38101
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-426, 'S', 428-486 <RES>
A:Cross-references: UNIPARC:UPI0000129E11; EMBL:X69438; NID:G38423; PIDN:CAA49214.1; PID
C:Genetics:
A:Gene: PAT133
A:Introns: 46/1

Query Match 85.7%; Score 36; DB 2; Length 486;
Best Local Similarity 87.5%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 279 GAEGLPGL 286

RESULT 13
T08613
hypothetical protein REA - Ehrlichia sp. (strain USG3)
C:Species: Ehrlichia sp.
A:Variety: strain USG3
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T08613
R:Storey, J.R.; Doros-Richert, L.A.; Gingrich-Baker, C.; Munroe, K.; Mather, T.N.; Cough
Infect. Immun. 66, 1356-1363, 1998
A:Title: Molecular cloning and sequencing of three granulocytic Ehrlichia genes encoding
A:Reference number: Z16457; MUID:98187902; PMID:9529053
A:Accession: T08613
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-619 <STO>
A:Cross-references: UNIPARC:UPI00000BC838; EMBL:AF020522; NID:G3033495; PID:G3033496
A:Experimental source: strain USG3
C:Genetics:
A:Gene: rea

Query Match 85.7%; Score 36; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AEGSPGL 8
||| |||
DB 15 AEGSPGL 21

RESULT 14
T22002
hypothetical protein F39H11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22002
```

```
R:White, S.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19500
A:Accession: T22002
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-650 <WIL>
A:Cross-references: UNIPROT:O17866; UNIPARC:UPI000006101E; EMBL:Z81079; PIDN:CA803084.1;
A:Experimental source: clone F39H11
C:Genetics:
A:Gene: CESP.F39H11.4
A:Map position: 1
A:Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 85.7%; Score 36; DB 2; Length 650;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 209 GADGAPGL 216

RESULT 15
A70812
hypothetical glycine-rich protein RV0833 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: A70812
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Suleston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70812
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-749 <COL>
A:Cross-references: UNIPROT:O53844; UNIPARC:UPI00000D4FD1; GB:AL022004; GB:AL123456; NI
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0833
C:Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containin

Query Match 85.7%; Score 36; DB 2; Length 749;
Best Local Similarity 87.5%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 740 GAEGLPGL 747

RESULT 16
S18251
collagen alpha 1(XI) chain - bovine (fragments)
C:Species: Bos primigenius taurus (cattle)
C>Date: 22-Nov-1993 #sequence_revision 23-Mar-1995 #text_change 09-Jul-2004
C:Accession: S18251; A46662; A56978; S65864; D46662; E46662; F46662; G46662; H46662; I
R:Brown, K.E.; Lawrence, R.; Sonenshein, G.E.
J. Biol. Chem. 266, 23268-23273, 1991
A:Title: Concerted modulation of alpha-1(XI) and alpha-2(V) collagen mRNAs in bovine va
A:Reference number: S18251; MUID:92078200; PMID:174123
A:Accession: S18251
A:Molecule type: mRNA
A:Residues: 1-911 <BRO>
A:Cross-references: UNIPROT:Q28083; UNIPARC:UPI0000126D29
R:Wayne, R.; Brewton, R.G.; Wayne, P.M.; Baker, J.R.
J. Biol. Chem. 268, 9381-9386, 1993
A:Title: Isolation and characterization of the chains of type V/type XI collagen presen
A:Reference number: A46662; MUID:93252802; PMID:8486632
```

A;Accession: C46662
A:Molecule type: protein
A;Residues: 347-354;356-363;586-600;912-924;925-961;962-998;999-1024 <MAY>
A;Cross-references: UNIPARC:UPI000017739D; UNIPARC:UPI000017739E; UNIPARC:UPI000017739F;
A;Experimental source: vitreous humor
A;Note: sequence modified after extraction from NCBI backbone
R;Wu, J. J.; Eyre, D. R.
J. Biol. Chem. 270, 18865-18870, 1995
A;Title: Structural analysis of cross-linking domains in cartilage type XI collagen. In
A;Reference number: A56978; MUID:95370194; PMID:7642541
A;Accession: A56978
A;Status: preliminary
A:Molecule type: protein
A;Residues: 206-229 <MAY>
A;Cross-references: UNIPARC:UPI00001773A4
R;Niyibizi, C.; Eyre, D. R.
Eur. J. Biochem. 224, 943-950, 1994
A;Title: Structural characteristics of cross-linking sites in type V collagen of bone. C
A;Reference number: S48210; MUID:95010086; PMID:7925418
A;Accession: S48210
A:Molecule type: protein
A;Residues: 'X', 273-298 <NIY>
A;Cross-references: UNIPARC:UPI00001773A5
A;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: hydroxyproline
F;211,223/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 85.7%; Score 36; DB 2; Length 1024;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|||:||||
DB 803 GAQGAPGL 810

RESULT 17
S23810
collagen alpha 1(XVI) chain precursor - human
N;Alternate names: procollagen alpha 1(XVI) chain
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text_change 09-Jul-2004
C;Accession: S23810; PQ0612; S08012
R;Pan, T.-C.; Zhang, R.-Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.
A;Reference number: S23810; MUID:92335339; PMID:1631157
A;Accession: S23810
A:Molecule type: mRNA
A;Residues: 1-1603 <PAN>
A;Cross-references: UNIPROT:Q07092; UNIPARC:UPI0000126D33; EMBL:M92642; NID:g180757; PID
A;Experimental source: skin fibroblasts
R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi
J. Biochem. 112, 856-863, 1992
A;Title: Molecular cloning and partial characterization of a novel collagen chain, alpha
A;Reference number: PQ0612; MUID:93203161; PMID:1284248
A;Accession: PQ0612
A:Molecule type: mRNA
A;Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1166-1603 <YAM>
A;Cross-references: UNIPARC:UPI000016B3C2; GB:S57132; NID:g298641; PIDN:AAB25797.1; PID:
A;Experimental source: Placenta
R;Kimura, S.
submitted to the EMBL Data Library, April 1989
A;Description: Partial nucleotide and amino acid sequence of a collagen-like protein fro
A;Reference number: S08012
A;Accession: S08012
A:Molecule type: mRNA
A;Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM' <KIM>
A;Cross-references: UNIPARC:UPI0000073DAB; EMBL:X14963; NID:g299984; PIDN:CAA33085.1; PID
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL16A1

A;Cross-references: GDB:I34045; OMIM:120326
A;Map position: lp34-lp34
C;Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI)
C;Function:
A;Description: structural component of extracellular fibrous polymer as a minor form pr
A;Note: may play a role in forming elastic connections at fibril surfaces
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxylysine
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>
F;22-333/Domain: amino-terminal nonhelical #status predicted <NC11>
F;334-1577/Region: interrupted helical
F;334-360/Domain: collagenous COL10 #status predicted <COL10>
F;375-505/Domain: collagenous COL9 #status predicted <COL9>
F;521-554/Domain: collagenous COL8 #status predicted <COL8>
F;539-541/Region: cell attachment (R-G-D) motif
F;572-630/Domain: collagenous COL7 #status predicted <COL7>
F;652-722/Domain: collagenous COL6 #status predicted <COL6>
F;738-875/Domain: collagenous COL5 #status predicted <COL5>
F;887-938/Domain: collagenous COL4 #status predicted <COL4>
F;973-987/Domain: collagenous COL3 #status predicted <COL3>
F;1005-1007/Region: cell attachment (R-G-D) motif
F;1011-1432/Domain: collagenous COL2 #status predicted <COL2>
F;1226-1228/Region: cell attachment (R-G-D) motif
F;1472-1577/Domain: collagenous COL1 #status predicted <COL1>
F;1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NCOL1>
F;47,327/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 85.7%; Score 36; DB 2; Length 1603;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|||:||||
DB 820 GAQGSFGV 827

RESULT 18
CGHUE
collagen alpha 1(XI) chain precursor - human
N;Alternate names: procollagen alpha 1(XI) chain
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence revision 03-Oct-1995 #text_change 08-May-1998
C;Accession: A35239; A31795
R;Yoshioka, H.; Ramirez, F.
J. Biol. Chem. 265, 6423-6426, 1990
A;Title: Pro-alpha(XI) collagen. Structure of the amino-terminal propeptide and express
A;Reference number: A35239; MUID:90202924; PMID:1690726
A;Accession: A35239
A:Molecule type: mRNA
A;Residues: 1-558 <YOS>
A;Cross-references: UNIPARC:UPI0000173BBF; GB:J05407
R;Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;
J. Biol. Chem. 263, 17159-17166, 1988
A;Title: Cloning and sequencing of pro-alpha(XI) collagen cDNA demonstrates that type XI
cartilaginous tissue.
A;Reference number: A92689; MUID:89034222; PMID:3182841
A;Accession: A31795
A:Molecule type: DNA; mRNA
A;Residues: 538-1806 <BER>
A;Cross-references: UNIPARC:UPI0000173BC0; GB:J04177
A;Note: parts of this sequence were determined by protein sequencing
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL11A1; COL16
A;Cross-references: GDB:120595; OMIM:120280
A;Map position: lp21-lp21
A;Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3
A;Note: the list of introns is incomplete
C;Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha
3(XI) chain (see PIR:CGHU6C), initially linked by disulfide bonds among their carboxyl-
rmed with desmosine cross-links made from lysine and allysine residues
C;Function:

A;Description: structural component of extracellular fibrous polymer associated with cell
A;Note: may play a role in controlling the lateral growth of collagen II fibrils
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-36/Domain: signal sequence #status predicted <SIG>
F;35-260/Domain: PARP-like #status predicted <PARP>
F;37-511/Domain: amino-terminal propeptide #status predicted <PRO>
F;512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>
F;512-527/Region: amino-terminal nonhelical telopeptide
F;528-1542/Region: helical
F;1543-1565/Region: carboxyl-terminal nonhelical telopeptide
F;1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>
F;1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;61-243.182-236/Diulfide bonds: #status predicted
F;505/Modified site: allysine (Lys) #status predicted
F;612.1452/Modified site: 5-hydroxylysine (Lys) #status predicted
F;612.1452/Binding site: carbohydrate (Lys) (covalent) #status predicted

Query Match 85.7%; Score 36; DB 1; Length 1806;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||:||||
DB 1036 GAQGAPGL 1043

RESULT 19
B89773
acetoin(diacetyl)reductase [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 05-Oct-2004
C;Accession: B89773
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Staphylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: B89773
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-258 <KUR>
A;Cross-references: UNIPROT:Q99X89; UNIPARC:UPI00000D7675; GB:BA000018; PID:g13700042; F
A;Experimental source: strain N315
C;Genetics:
A;Gene: butA
C;Superfamily: short-chain dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 83.3%; Score 35; DB 2; Length 258;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||:||||
DB 144 GVEGNFGL 151

RESULT 20
JC1448
collagen col-34 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: JC1448
R;Bird, D.M.
Gene 120, 261-266, 1992
A;Title: Sequence comparison of the Caenorhabditis elegans dpy-13 and col-34 genes, and
A;Reference number: JC1448; MUID:93013043; PMID:1398138
A;Accession: JC1448
A;Molecule type: DNA
A;Residues: 1-298 <BIR>
A;Cross-references: UNIPROT:P34687; UNIPARC:UPI000016B8EC; GB:M80650; NID:g156249; PIDN:
C;Genetics:

A;Gene: col-34

Query Match 83.3%; Score 35; DB 2; Length 298;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
||:||||
DB 236 GADGSPG 242

RESULT 21

T29956

hypotheical protein F36A4.10 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T29956

R;Miller, N.; Bradshaw, H.

submitted to the EMBL Data Library, April 1996

A;Description: The sequence of C. elegans cosmid F36A4.

A;Reference number: Z20713

A;Accession: T29956

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-299 <MIL>

A;Cross-references: UNIPROT:P34687; UNIPARC:UPI0000127237; EMBL:U53333; PIDN:AAA96155.1.

A;Experimental source: strain Bristol N2; clone F36A4

C;Genetics:

A;Gene: CESP:F36A4.10

A;Map position: 4

A;Introns: 27/3; 243/1

Query Match

83.3%; Score 35; DB 2; Length 299;

Best Local Similarity 85.7%; Pred. No. 49;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7

||:||||

DB 237 GADGSPG 243

RESULT 22

T32247

hypotheical protein T15B7.5 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C;Accession: T32247

R;Pauley, A.; Gattung, S.

submitted to the EMBL Data Library, September 1997

A;Description: The sequence of C. elegans cosmid T15B7.

A;Reference number: Z21139

A;Accession: T32247

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-314 <PAU>

A;Cross-references: UNIPROT:O17035; UNIPARC:UPI00000811BD; EMBL:AF022985; PIDN:AAB69960

A;Experimental source: strain Bristol N2; clone T15B7

C;Genetics:

A;Gene: CESP:T15B7.5

A;Map position: 5

A;Introns: 273/1

Query Match

83.3%; Score 35; DB 2; Length 314;

Best Local Similarity 75.0%; Pred. No. 51;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8

||:||||

DB 108 GAQGEPL 115

RESULT 23

S02170

collagen alpha 1(IX) chain - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 09-Jul-2004
C;Accession: S02170
R;Kimura, T.; Mattei, M.G.; Stevens, J.W.; Goldring, M.B.; Ninomiya, Y.; Olsen, B.R.
Eur. J. Biochem. 179, 71-78, 1989
A;Title: Molecular cloning of rat and human type IX collagen cDNA and localization of the
A;Reference number: S02140; MUID:89137096; PMID:2465149
A;Accession: S02170
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-325 <KIN>
A;Cross-references: UNIPROT:P20850; UNIPARC:UPI0000126D25
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer

Query Match 83.3%; Score 35; DB 2; Length 325;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEKSPGL 8
| | | | |
Db 35 GPEGSPGI 42

RESULT 24
T26281
hypothetical protein W08D2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T26281
R;Swinburne, J.; Ainscough, R.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z20188
A;Accession: T26281
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-358 <WIL>
A;Cross-references: UNIPROT:Q23222; UNIPARC:UPI000007AB54; EMBL:Z70271; PIDN:CAA94234.1;
A;Experimental source: clone W08D2
C;Genetics:
A;Gene: CESP:W08D2.6
A;Map position: 4
A;Introns: 57/3; 320/2

Query Match 83.3%; Score 35; DB 2; Length 358;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEKSPGL 8
| | | | |
Db 149 GAKGAPGL 156

RESULT 25
I50629
collagen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I50629
R;Fuller, F.; Boedker, H.
Biochemistry 20, 996-1006, 1981
A;Title: Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I)
A;Reference number: I50623; MUID:81160715; PMID:6927845
A;Accession: I50629
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-473 <FUL>
A;Cross-references: UNIPROT:P02457; UNIPARC:UPI00001712E8; EMBL:V00401; NID:G63307; PIDN:
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;244-473/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 83.3%; Score 35; DB 2; Length 473;
Best Local Similarity 85.7%; Pred. No. 78;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEKSPG 7
| | | | |
Db 28 GAEKAPG 34

RESULT 26
S41067
collagen alpha 1(III) chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S41067; A29905; S31924
R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.
Biochim. Biophys. Acta 1217, 41-48, 1994
A;Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa
A;Reference number: S41067; MUID:94114571; PMID:8286415
A;Accession: S41067
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-636 <GLU>
A;Cross-references: UNIPROT:P13941; UNIPARC:UPI0000126D1A; EMBL:X70369; NID:G57915; PIDN:
R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Komm, B.; Mohn, K.
DNA 7, 347-354, 1988
A;Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2 (V) collagen mRNAs by estr
A;Reference number: A29905; MUID:88296083; PMID:2456904
A;Accession: A29905
A;Molecule type: mRNA
A;Residues: 308-482 <FRA>
A;Cross-references: UNIPARC:UPI0000170BC0; GB:M21354; NID:G203500; PIDN:AAA40942.1; PID:
R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.
submitted to the EMBL Data Library, February 1993
A;Reference number: S31924
A;Accession: S31924
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 2-636 <GL2>
A;Cross-references: UNIPARC:UPI000017737E; EMBL:X70369
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 83.3%; Score 35; DB 2; Length 636;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEKSPG 7
| | | | |
Db 345 GSEKSPG 351

RESULT 27
S20819
collagen alpha 3(IX) chain precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004
C;Accession: S20819; S22429; S22918; S22238; C18856; S22241
R;Brewton, R.G.; Ouspenskaia, M.V.; van der Rest, M.; Mayne, R.
Eur. J. Biochem. 205, 443-449, 1992
A;Title: Cloning of the chicken alpha-3(IX) collagen chain completes the primary structu
A;Reference number: S20819; MUID:92241276; PMID:1572350
A;Accession: S20819
A;Molecule type: mRNA
A;Residues: 1-675 <BRE>
A;Cross-references: UNIPROT:Q90800; UNIPARC:UPI00000FDCD2; EMBL:X64712; NID:G63316; PIDN:
R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;
J. Biol. Chem. 267, 10070-10076, 1992
A;Title: Cloning and developmental expression of the alpha3 chain of chicken type IX col
A;Reference number: S22429; MUID:92250566; PMID:1577778
A;Accession: S22429
A;Molecule type: mRNA
A;Residues: 1-195, 'G', 197-675 <HAL>
A;Cross-references: UNIPARC:UPI000017A144; EMBL:M83179

A;Note: 353-Arg, 386-Leu and 548-Arg were also found
 R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Uyeyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;
 submitted to the EMBL Data Library, February 1992
 A;Description: Cloning and developmental expression of the alpha 3 chain of chicken type
 A;Reference number: S22918
 A;Accession: S22918
 A;Molecule type: mRNA
 A;Residues: 1-195, 'G', 197-405, 'S', 407-675 <HA2>
 A;Cross-references: UNIPARC:UPI0000126D50; EMBL:M83179; NID:g211040; PIDN:AAB59960.1; PI
 R;Mayne, R.; van der Rest, M.; Nimomiya, Y.; Olsen, B.R.
 Ann. N. Y. Acad. Sci. 460, 38-46, 1985
 A;Title: The structure of type IX collagen.
 A;Reference number: S22238; MUID:86185164; PMID:3868958
 A;Accession: S22238
 A;Molecule type: protein
 A;Residues: 540-548 <MAY>
 A;Cross-references: UNIPARC:UPI000017A145
 R;Nimomiya, Y.; van der Rest, M.; Mayne, R.; Lozano, G.; Olsen, B.R.
 Biochemistry 24, 4223-4229, 1985
 A;Title: Construction and characterization of cDNA encoding the alpha2 chain of chicken
 A;Reference number: A18856; MUID:86026268; PMID:2996593
 A;Accession: C18856
 A;Molecule type: protein
 A;Residues: 540-558 <NIN>
 A;Cross-references: UNIPARC:UPI000017A146
 R;Shimokomaki, M.; Wright, D.W.; Irwin, M.H.; van der Rest, M.; Mayne, R.
 Ann. N. Y. Acad. Sci. 580, 1-7, 1990
 A;Title: The structure and macromolecular organization of type IX collagen in cartilage.
 A;Reference number: S22241; MUID:90347791; PMID:2186687
 A;Accession: S22241
 A;Molecule type: protein
 A;Residues: 135, 'E', 137-187, 'X', 189-191 <SHI>
 A;Cross-references: UNIPARC:UPI000017A147
 C;Keywords: coiled coil; connective tissue; disulfide bond; extracellular matrix; hetero
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-675/Product: collagen alpha 3(IX) chain #status predicted <MAT>
 F;22-24/Domain: non-collagenous NC4 #status predicted <NC4>
 F;25-161/Domain: non-collagenous COL3 #status predicted <COL3>
 F;162-176/Domain: non-collagenous NC3 #status predicted <NC3>
 F;177-515/Domain: collagenous COL2 #status predicted <COL2>
 F;516-546/Domain: non-collagenous NC2 #status predicted <NC2>
 F;547-657/Domain: collagenous COL1 #status predicted <COL1>
 F;658-675/Domain: non-collagenous NC1 #status predicted <NC1>
 F;22/Modified site: pyrrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F;137,143,146,149,152,155,179,182,185,552,558/Modified site: hydroxyproline (Pro) #statu
 F;170,174,525,658,663/Disulfide bonds: interchain #status predicted

Query Match 83.3%; Score 35; DB 2; Length 675;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 Db 632 GAEGSPGL 639

RESULT 28
 EDBEXD
 immediate-early protein RL2 - human herpesvirus 2 (strain HG52)
 N;Alternate names: RL2 protein
 C;Species: human herpesvirus 2
 A;Note: host Homo sapiens (man)
 C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
 C;Accession: JQ1501
 R;McGeoch, D.J.; Cunningham, C.; McIntyre, G.; Dolan, A.
 J. Gen. Virol. 72, 3057-3075, 1991
 A;Title: Comparative sequence analysis of the long repeat regions and adjoining parts of
 A;Reference number: JQ1494; MUID:92113549; PMID:1662697
 A;Accession: JQ1501
 A;Molecule type: DNA
 A;Residues: 1-825 <MCG>
 A;Cross-references: UNIPROT:P28284; UNIPARC:UPI000012D179; GB:D10471; DDBJ:D01128; NID:9
 C;Genetics:

A;Gene: RL2
 A;Intons: 25/3; 252/1
 C;Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
 C;Keywords: DNA binding; immediate-early protein; tandem repeat; transcription regulation
 F;122-172/Domain: RING finger homology <RNG>
 F;128-166/Region: zinc finger C3HC4 motif
 F;589-623/Region: 5-residue repeats (A-S-S-S-S)

Query Match 83.3%; Score 35; DB 1; Length 825;
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 Db 484 GAEGSPGV 491

RESULT 29
 S28791
 collagen alpha 1(XI) chain - chicken (fragment)
 C;Species: Gallus gallus (chicken)
 C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
 C;Accession: S28791
 R;Nah, H.D.; Barembaum, M.; Upholt, W.B.
 J. Biol. Chem. 267, 22581-22586, 1992
 A;Title: The chicken alpha1(XI) collagen gene is widely expressed in embryonic tissues.
 A;Reference number: S28791; MUID:93054557; PMID:1429607
 A;Accession: S28791
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-888 <NAH>
 A;Cross-references: UNIPROT:Q0796; UNIPARC:UPI00000FBAD3; EMBL:M88593; NID:g211619; PI
 C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
 F;665-887/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 83.3%; Score 35; DB 2; Length 888;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
 |||||
 Db 466 GAEGAPG 472

RESULT 30
 S40495
 collagen alpha 1(IX) chain, long form - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
 C;Accession: S40495
 R;Abe, N.; Yoshioka, H.; Inoue, H.; Nimomiya, Y.
 Biochim. Biophys. Acta 1204, 61-67, 1994
 A;Title: The complete primary structure of the long form of mouse alpha-1(IX) collagen.
 A;Reference number: S40495; MUID:94137762; PMID:8305476
 A;Accession: S40495
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-921 <ABE>
 A;Cross-references: UNIPROT:Q05722; UNIPARC:UPI00000278C0; GB:D17511; NID:g511661; PIDN

Query Match 83.3%; Score 35; DB 2; Length 921;
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 Db 631 GPEGSPGI 638

RESULT 31
 S42617
 collagen alpha 1(IX) chain - mouse
 C;Species: Mus musculus (house mouse)

C;Date: 25-Dec-1994 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S42617
R;Kokos, I.; Muragaki, Y.; Warman, M.; Oleen, B.R.
Matrix Biol. 14, 1-8, 1994
A;Title: Assembly and sequencing of a cDNA covering the entire mouse alpha-1(IX) collage
A;Reference number: S42617; MUID:94340199; PMID:8061915
A;Accession: S42617
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-921 <RQ>
A;Cross-references: UNIPROT:Q05722; UNIPARC:UPI000016CCB7; GB:L12215; GB:L19396; NID:g19

Query Match 83.3%; Score 35; DB 2; Length 921;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
| | | | |
Db 631 GPEGSPGI 638

RESULT 32
CGCH1S
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)
C;Species: Gallus gallus (chicken)
C;Date: 12-Aug-1981 #sequence_revision 06-Jul-1982 #text_change 31-Mar-2000
C;Accession: A90458; A90181; A02857
R;Higberger, J.H.; Corbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross, J.
Biochemistry 21, 2048-2055, 1982
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-CB8 and the complete prima
A;Reference number: A90458; MUID:82231995; PMID:7093229
A;Accession: A90458
A;Molecule type: protein
A;Residues: 1-1036 <HIG>
A;Cross-references: UNIPARC:UPI0000173B62
A;Experimental source: skin
A;Note: This is the latest in a series of papers from these workers elucidating the sequ
R;Eyre, D.R.; Glimcher, M.J.
Biochem. Biophys. Res. Commun. 48, 720-726, 1972
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the alp
A;Reference number: A90181; MUID:72243016; PMID:5047697
A;Accession: A90181
A;Molecule type: protein
A;Residues: 1037-1042 <EVR>
A;Cross-references: UNIPARC:UPI0000173B63
A;Experimental source: skin
A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein
C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some
C;Comment: Most of the prolines at the third position of the tripeptide repeating unit
C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in po
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;
F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match 83.3%; Score 35; DB 1; Length 1042;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 857 GAEGAPG 863

RESULT 33
CGB07S
collagen alpha 1(III) chain - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1986 #sequence_revision 04-Dec-1986 #text_change 09-Jul-2004
C;Accession: A02862; A38001; A38002; A38004; A38005; S71946
R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wachter, E.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979
A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen
A;Reference number: A02862; MUID:80026026; PMID:488906

A;Accession: A02862
A;Molecule type: protein
A;Residues: 1-242 <FIE>
A;Cross-references: UNIPROT:P04258; UNIPARC:UPI0000173B8A
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979
A;Title: The covalent structure of calf skin type III collagen. II. The amino acid sequ
A;Reference number: A38001; MUID:80026027; PMID:488907
A;Accession: A38001
A;Molecule type: protein
A;Residues: 243-422 <DEW1>
A;Cross-references: UNIPARC:UPI0000173B8B
R;Bentz, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979
A;Title: The covalent structure of calf skin type III collagen. III. The amino acid sequ
A;Reference number: A38002; MUID:80026028; PMID:488908
A;Accession: A38002
A;Molecule type: protein
A;Residues: 423-571 <BEN>
A;Cross-references: UNIPARC:UPI0000173B8C
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequ
A;Reference number: A38003; MUID:80026029; PMID:488909
A;Accession: A38003
A;Molecule type: protein
A;Residues: 572-808 <LAN>
A;Cross-references: UNIPARC:UPI0000173B8D
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid sequen
A;Reference number: A38004; MUID:80026030; PMID:488910
A;Accession: A38004
A;Molecule type: protein
A;Residues: 809-947 <DEW2>
A;Cross-references: UNIPARC:UPI0000173B8E
R;Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-869, 1979
A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequ
A;Reference number: A38005; MUID:80026031; PMID:488911
A;Accession: A38005
A;Molecule type: protein
A;Residues: 948-1049 <ALL>
A;Cross-references: UNIPARC:UPI0000173B8F
A;Experimental source: skin
R;Henkel, W.
Biochem. J. 318, 497-503, 1996
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
A;Reference number: S71946; MUID:96404897; PMID:8809038
A;Accession: S71946
A;Molecule type: protein
A;Residues: 87-106;1017-1029;1037-1049 <HEN>
A;Cross-references: UNIPARC:UPI0000173B90; UNIPARC:UPI0000173B91; UNIPARC:UPI0000173B92
C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are h
C;Comment: The type III collagen molecule is a trimer of identical chains, linked to eac
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CAB>
F;1-14/Region: amino-terminal nonhelical telopeptide
F;15-1040/Region: helical
F;587-589/Region: cell attachment (R-G-D) motif
F;752-754/Region: cell attachment (R-G-D) motif
F;875-877/Region: cell attachment (R-G-D) motif
F;878-880/Region: cell attachment (R-G-D) motif
F;935-937/Region: cell attachment (R-G-D) motif
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F;95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
F;107,950/Modified site: allysine (Lys) #status predicted
F;107/Binding site: carbonylate (Lys) (covalent) #status experimental
F;1040,1041/Diulfide bonds: interchain #status predicted

Query Match 83.3%; Score 35; DB 1; Length 1049;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|:|||||
Db 1017 GSEGSFG 1023

RESULT 34

A70934
Hypothetical glycine-rich protein RV0578c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: A70934
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70934
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1306 <COL>
A:Cross-references: UNIPROT:O53775; UNIPARC:UPI00000D4FC8; GB:AL021942; GB:AL123456; NID:
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV0578c
C:Superfamily: collagen alpha 1(IV) chain

Query Match 83.3%; Score 35; DB 2; Length 1306;
Best Local Similarity 85.7%; Pred No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|:|||||
Db 118 GADGSPG 124

RESULT 35

CGHU7L
collagen alpha 1(III) chain precursor - human
N:Alternate names: procollagen alpha 1(III) chain
C:Species: Homo sapiens (nan)
C:Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399
R:Prockop, D.J.
submitted to the EMBL Data Library, February 1989
A:Reference number: S05272
A:Accession: S05272
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1240, 'V', 1242-1466 <PRC>
A:Cross-references: UNIPROT:P02461; UNIPARC:UPI0000000CDE; EMBL:X14420; NID:g30057; PIDN:
R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
Biochem. J. 260, 509-516, 1989
A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human
erences.
A:Reference number: S04642; MUID:89350838; PMID:2764886
A:Accession: S04642
A:Molecule type: mRNA
A:Residues: 1-1196 <ALA>
A:Cross-references: UNIPARC:UPI0000173B80; EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:
A:Note: the complete sequence is not shown
R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
Gene 78, 255-265, 1989
A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene
A:Reference number: PE0011; MUID:89378752; PMID:2777083
A:Accession: PE0011
A:Molecule type: DNA
A:Residues: 1-176 <BEN>
A:Cross-references: UNIPARC:UPI000016A703; GB:M26939; NID:g180813; PIDN:AAAS2040.1; PID:
R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.

Nucleic Acids Res. 16, 7201, 1988
A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre
A:Reference number: S01726; MUID:88303360; PMID:3405773
A:Accession: S01726
A:Molecule type: mRNA
A:Residues: 1-170 <TOM>
A:Cross-references: UNIPARC:UPI000016A706; EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:
A:Note: the authors translated the codon CAG for residue 154 as His
R:Janeczko, R.A.; Ramirez, F.
Nucleic Acids Res. 17, 6742, 1989
A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen
A:Reference number: S04887; MUID:89386015; PMID:2780304
A:Accession: S04887
A:Molecule type: mRNA
A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A:Cross-references: UNIPARC:UPI000016A61C; EMBL:X15332; NID:g29545; PIDN:CAA33387.1; PID:
A:Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R:Seyer, J.M.; Kang, A.H.
Biochemistry 16, 1158-1164, 1977
A:Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A:Reference number: A90399; MUID:77134724; PMID:557335
A:Accession: A90399
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A:Cross-references: UNIPARC:UPI0000173B81
A:Experimental source: liver
A:Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galac-
R:Seyer, J.M.
submitted to the Atlas, December 1977
A:Reference number: A94562
A:Accession: A94562
A:Molecule type: protein
A:Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A:Cross-references: UNIPARC:UPI0000173B82
A:Experimental source: liver
A:Note: author submitted corrections to A90399
R:Miliewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A:Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusual
splicing.
A:Reference number: I51868; MUID:93304430; PMID:8317500
A:Accession: I51868
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 186-194 <MIL>
A:Cross-references: UNIPARC:UPI0000000B14; GB:S62925; NID:g386425; PIDN:AAD13937.1; PID:
R:Chodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A:Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL
A:Reference number: S59511; MUID:96067614; PMID:7487954
A:Accession: S59511
A:Molecule type: mRNA
A:Residues: 302-423 <CHI>
A:Cross-references: UNIPARC:UPI0000173B83; GB:S79877; NID:g1195576; PIDN:AAB35615.1; PID:
R:Seyer, J.M.; Kang, A.H.
Biochemistry 17, 3404-3411, 1978
A:Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr p
A:Reference number: A90414; MUID:79000343; PMID:687591
A:Accession: A90414
A:Molecule type: protein
A:Residues: 399-675, 'N', 677-727 <SEY3>
A:Cross-references: UNIPARC:UPI0000173B84
A:Experimental source: liver
R:Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A:Title: G to T transversion at position +5 of a splice donor site causes skipping of ti
A:Reference number: I55349; MUID:91161621; PMID:1672129
A:Accession: I55349
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 537-605 <LEE>
A:Cross-references: UNIPARC:UPI00000000AA2; GB:M59312; NID:g180815; PIDN:AAAS2041.1; PID:
R:Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from ty
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895,'A',897-964 <SEY4>
A;Cross-references: UNIPARC:UPI0000173B85
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Laman, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping an
A;Reference number: A38303; MUID:91009133; PMID:2145268
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: UNIPARC:UPI000000004A1; GB:J05617; GB:M55603; GB:M59227; NID:g180878;
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syn
R;Mankoo, B.S.; Dalgleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>
A;Cross-references: UNIPARC:UPI0000173B86; EMBL:X06700; NID:g30053; PIDN:CAA29886.1; PID
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979,'A',981-984,'PS',987,'QN',990-1096,'P',1098-1152,'AT',1155,'S',1157-
A;Cross-references: UNIPARC:UPI0000173B87
A;Experimental source: liver
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Mye
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollage
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155,'P',1157-1466 <LOI>
A;Cross-references: UNIPARC:UPI0000173B88; EMBL:X01655; EMBL:X01742; NID:g29584; PIDN:CA
R;Miskulin, M.; Dalgleish, R.; Kluge-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 23, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: UNIPARC:UPI000016AGB5; GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3858826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: UNIPARC:UPI000016AGB6; GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHU>
A;Cross-references: UNIPARC:UPI0000173B89; GB:M10615; GB:M10793; GB:M10794; GB:M10795; C
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
action
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit

3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently O
C;Genetics:
A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide bo
er of their length, is formed with desmosine cross-links made from lysine and allysine r
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inter
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F:31-91/Domain: von Willebrand factor type C repeat homology <VMC>
F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F:154-167/Region: amino-terminal nonhelical telopeptide
F:168-1196/Region: helical
F:1091-1093/Region: cell attachment (R-G-D) motif
F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F:161,1212/Modified site: allysine (Lys) #status predicted
Query Match 83.3%; Score 35; DB 1; Length 1466;
Best Local Similarity 85.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPG 7
Db 1173 GSEGSPPG 1179
RESULT 36
CGHU3B
collagen alpha 3(IV) chain precursor, long splice form - human
N;Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice form
C;Species: Homo sapiens (man)
C;Date: 28-Oct-1994 #sequence revision 03-Oct-1995 #text_change 09-Jul-2004
C;Accession: A54763; A43928; A44043; A45971; A39786
R;Maruyama, M.; Iainonen, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A;Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpression
A;Reference number: A54763; MUID:94364994; PMID:8083201
A;Accession: A54763
A;Molecule type: mRNA
A;Residues: 1-1670 <VAR>
A;Cross-references: UNIPROT:Q01955; UNIPARC:UPI0000173BE3; GB:X80031; NID:g577563; PID:g
A;Experimental source: kidney
R;Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A;Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the al
A;Reference number: A43928; MUID:92147878; PMID:1737849
A;Accession: A43928
A;Molecule type: mRNA
A;Residues: 1331-1524,'I',1526-1670 <TUR>
A;Cross-references: UNIPARC:UPI0000173BE4; GB:M81379
A;Experimental source: kidney
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A;Title: Exon/Intron structure of the human alpha 3(IV) gene encompassing the Goodpastur
ction.
A;Reference number: A44043; MUID:93015826; PMID:1400291
A;Accession: A44043
A;Molecule type: DNA
A;Residues: 1386-1670 <QUI>
A;Cross-references: UNIPARC:UPI000016A42D; GB:M29293; NID:g177895; PIDN:AAA21610.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIP:115597)
R;Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994
 A:Reference number: A44738; MUID:94274734; PMID:8006044
 A:Contents: annotation; erratum; correction to intronic sequence in A44043
 R:Bernal, D.; Quinones, S.; Saub, J.
 J. Biol. Chem. 268, 12090-12094, 1993
 A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
 A:Reference number: A45971; MUID:93280184; PMID:8505332
 A:Accession: A45971
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1427-1444 <BER>
 A:Cross-references: UNIPARC:UPI0000173BES
 A:Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly identified
 R:Morrison, K.E.; Mariyama, M.; Yang-Peng, T.L.; Redders, S.T.
 Am. J. Hum. Genet. 49, 545-554, 1991
 A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain of
 A:Reference number: A39786; MUID:91353570; PMID:1882840
 A:Accession: A39786
 A:Molecule type: mRNA
 A:Residues: 1453-1593 'A', 1595-1670 <MOR>
 A:Cross-references: UNIPARC:UPI000014C40B; GB:S55790; NID:q234418; PIDN:AAB19637.1; PID:
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C:Genetics:
 A:Gene: GDB:COL4A3
 A:Cross-references: GDB:128351; OMIM:120070
 A:Map position: 2q36-q37
 A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
 A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
 er associations in the interrupted helical domain (with disulfide and desmosine cross-l
 C:Function:
 A:Description: minor structural component of extracellular basement membrane in kidney g
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 F:1-28/Domain: signal sequence #status predicted <SIG>
 F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
 F:29-42/Domain: amino-terminal nonhelical, NH1 <NHI>
 F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:31.33, 39, 41, 125, 422, 476, 479, 682, 809, 1387/Disulfide bonds: interchain #status predi
 F:253/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:1460-1548/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511, 1616-1622/Disulfide bonds: #status predicted
 F:1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 83.3%; Score 35; DB 1; Length 1670;
 Best Local Similarity 75.0%; Pred. No. 2.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GAEGSPGL 8
 DB 529 GAQGDFGL 536
 RESULT 37
 A45407
 collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)
 C:Species: Strongylocentrotus purpuratus (purple urchin)
 C:Date: 22-Sep-1993 #sequence revision 18-Nov-1994 #text_change 31-Dec-2004
 C:Accession: A45407; A43903; A23940
 R:Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
 J. Biol. Chem. 268, 5249-5254, 1993

A:Title: Complete primary structure of a sea urchin type IV collagen alpha chain and ana
 A:Reference number: A45407; MUID:93186842; PMID:8444899
 A:Accession: A45407
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-1752 <EXP>
 A:Cross-references: UNIPROT:Q26312; UNIPARC:UPI00001773DE
 A:Note: sequence extracted from NCBI backbone (NCBI:126841)
 R:Wessel, G.M.; Etkin, M.; Benson, S.
 Dev. Biol. 148, 261-272, 1991
 A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously produ
 A:Reference number: A43903; MUID:92038439; PMID:1936564
 A:Accession: A43903
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'P', 633-1537, 'G' <WES>
 A:Cross-references: UNIPARC:UPI000007C802; GB:S64572; NID:q238616; PIDN:AAB20270.1; PID
 A:Note: sequence extracted from NCBI backbone (NCBI:64572, NCBIP:64573)
 R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.
 Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
 A:Title: Structure and developmentally regulated expression of a Strongylocentrotus purp
 A:Reference number: A23940; MUID:86205894; PMID:3458186
 A:Accession: A23940
 A:Molecule type: DNA
 A:Residues: 742-812 <VEN>
 A:Cross-references: UNIPARC:UPI00001773DF; EMBL:M13206
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F:29-163/Domain: amino-terminal nonhelical, 7S <7SD>
 F:162-1523/Region: interrupted helical
 F:1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F:129/Modified site: allysine (Lys) #status predicted

Query Match 83.3%; Score 35; DB 2; Length 1752;
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPGL 8
 DB 1189 GAGNPGGL 1196
 RESULT 38
 T30258
 adenomatous polyposis coli protein 2 - mouse
 N:Alternate names: APC2 protein
 C:Species: Mus musculus (house mouse)
 C:Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T30258
 R:van Es, J.H.; Kirkpatrick, C.; van de Wetering, M.; Molenaar, M.; Kuipers,
 Curr. Biol. 9, 105-108, 1999
 A:Title: Identification of APC2, a homologue of the adenomatous polyposis coli tumour a
 A:Reference number: Z20796; MUID:99147086; PMID:10021369
 A:Accession: T30258
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2274 <VAN>
 A:Cross-references: UNIPROT:Q9ZIK7; UNIPARC:UPI00000296CC; EMBL:AJ130783; NID:94210431;
 C:Genetics:
 A:Gene: APC2
 A:Introns: 47/3; 78/1; 138/2; 174/3; 212/3; 238/3; 271/3; 396/1; 428/1; 474/3; 500/3; 5
 Query Match 83.3%; Score 35; DB 2; Length 2274;
 Best Local Similarity 85.7%; Pred. No. 3.8e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPGL 7
 DB 1929 GAEGTNG 1935
 RESULT 39

A75508
hypothetical protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: A75508
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; MUID:20036896; PMID:10567266
A/Accession: A75508
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-206 <WHI>
A/Cross-references: UNIPROT:Q9RWY4; UNIPARC:UPI00000C178A; GB:AE001911; GB:AE000513; NID
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0531
A/Map position: 1
C/Superfamily: Deinococcus radiodurans hypothetical protein DR0531

Query Match 81.0%; Score 34; DB 2; Length 206;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 114 GGEGLSPG 120
| | | | |

RESULT 40
S42627
glycine N-methyltransferase (EC 2.1.1.20) - human
N/Alternate names: folate-binding protein
C/Species: Homo sapiens (man)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Dec-2000
C/Accession: S42627; S44383
R/Ogawa, H.; Gomi, T.; Fujioka, M.
Comp. Biochem. Physiol. B 106, 601-611, 1993
A/Title: Mammalian glycine N-methyltransferases. Comparative kinetic and structural prop
A/Reference number: S42627; MUID:94109127; PMID:8281755
A/Accession: S42627
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-295 <OGA>
A/Cross-references: UNIPARC:UPI00001722F8; EMBL:X62250; NID:G433939
R/Ogawa, H.
submitted to the EMBL Data Library, September 1991
A/Reference number: S44383
A/Accession: S44383
A/Molecule type: mRNA
A/Residues: 1-118,120-174,'S',176-255,'S','PSS',261-295 <OG2>
A/Cross-references: UNIPARC:UPI00001722F9; EMBL:X62250; NID:G433939; PIDN:CAA44164.1; PI
A/Note: the translated sequence in GenBank entry HSLGMEFF, release 114.0, (PIDN:AA44164.
C/Superfamily: glycine N-methyltransferase
C/Keywords: acetylated amino end; folate; methyltransferase; S-adenosylmethionine
F:2-295/Product: glycine N-methyltransferase #status predicted <MAT>
F:2/Modified site: acetylated amino end (Val) (in mature form) #status predicted

Query Match 81.0%; Score 34; DB 1; Length 295;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 231 GQDGLSPGL 238
| | | | |

RESULT 41
S55900
DNAJ-like protein homolog - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe

C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C/Accession: S55900
R/Park, S.K.; Chon, S.K.; Yoo, H.S.
Biochim. Biophys. Acta 1262, 87-90, 1995
A/Title: A cDNA of Schizosaccharomyces pombe encoding a homologue of DnaJ-like protein.
A/Reference number: S55900; MUID:95290501; PMID:7772606
A/Accession: S55900
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-379 <PAR>
A/Cross-references: UNIPROT:Q09912; UNIPARC:UPI000001691AD; EMBL:L37753; NID:G576932; PID
C/Genetics:
A/Gene: psi
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 81.0%; Score 34; DB 2; Length 379;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 80 GAEGGPG 86
| | | | |

RESULT 42
T41633
psi protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T41633
R/McDougall, R.M.; Rajandream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.
submitted to the EMBL Data Library, August 1999
A/Reference number: Z22005
A/Accession: T41633
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-379 <MCD>
A/Cross-references: UNIPROT:Q09912; UNIPARC:UPI000001327AF; EMBL:AL109850; PIDN:CAB52880.
A/Experimental source: strain 972h-; cosmid c830
C/Genetics:
A/Gene: SPDB:SPCC830.07C
A/Map position: 3
C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 81.0%; Score 34; DB 2; Length 379;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 80 GAEGGPG 86
| | | | |

RESULT 43
T31115
histidine kinase homolog sdek - Myxococcus xanthus
C/Species: Myxococcus xanthus
C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C/Accession: T31115
R/Garza, A.G.; Pollack, J.S.; Harris, B.Z.; Lee, A.; Keseler, I.; Licking, E.F.; Singer,
submitted to the EMBL Data Library, October 1997
A/Description: A histidine kinase is required early in fruiting body development in myxo
A/Reference number: Z20990
A/Accession: T31115
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-513 <GAR>
A/Cross-references: UNIPROT:O52237; UNIPARC:UPI00000AF046; EMBL:AF031084; NID:G2736295;
C/Genetics:
A/Gene: sdek

Query Match 81.0%; Score 34; DB 2; Length 513;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
||| |||
DB 225 GAEGGPG 231

RESULT 44
S32436
collagen alpha 2(IX) chain - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Sep-2003
C:Accession: S32436; S34487; S64673
R:Perlaelae, M.; Haenninen, M.; Haestbacka, J.; Elima, K.; Vuorio, E.
FEBS Lett. 319, 177-180, 1993
A>Title: Molecular cloning of the human alpha-2(IX) collagen cDNA and assignment of the
A:Reference number: S32436; MUID:93202262; PMID:8454052
A:Accession: S32436
A:Molecule type: mRNA
A:Residues: 1-618 <PER1>
A:Cross-references: UNIPARC:UPI000017A166; EMBL:M95610; NID:g1054872
R:Perlaelae, M.; Haenninen, M.; Haestbacka, J.; Vuorio, E.
submitted to the EMBL Data Library, March 1993
A:Description: Molecular cloning of the human alpha-2 (IX) collagen cDNA and assignment
A:Reference number: S34487
A:Accession: S34487
A:Molecule type: mRNA
A:Residues: 1-26, 'QT', 29, 'S', 31-32, 'IM', 35-561, 'L', 563-578, 'P', 580-618 <PER2>
A:Cross-references: UNIPARC:UPI000017A167; EMBL:M95610; NID:g1054872
R:Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A>Title: Collagen type IX from human cartilage: a structural profile of intermolecular
A:Reference number: S64673; MUID:96195147; PMID:8660302
A:Accession: S64673
A:Molecule type: protein
A:Residues: 123-133, 'P', 135-137 <DIA>
A:Cross-references: UNIPARC:UPI000017A168
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL9A2
A:Cross-references: GDB:138310; OMIM:120260
A:Map position: lp33-lp32.2
C:Complex: type IX collagen may be a heterotrimer of one alpha 1(IX) chain, one alpha 2
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with type
A:Keywords: chondroitin sulfate proteoglycan; coiled coil; extracellular matrix; glycop
F;1-114/Domain: collagenous COL3 (fragment) #status predicted <COL3>
F;115-131/Domain: non-collagenous NC3 #status predicted <NC3>
F;132-470/Domain: collagenous COL2 #status predicted <COL2>
F;471-500/Domain: non-collagenous NC2 #status predicted <NC2>
F;501-615/Domain: collagenous COL1 #status predicted <COL1>
F;616-618/Domain: non-collagenous NC1 (fragment) #status predicted <NC1>
F;120/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 81.0%; Score 34; DB 2; Length 618;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 231 GDEGSPGI 238

RESULT 45
A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C>Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C:Accession: A55267
R;Zheng, K.; Thorner, P.S.; Marrano, P.; Baumal, R.; McInnes, R.R.

Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A>Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-l
en type IV.
A:Reference number: A55267; MUID:94224868; PMID:8171024
A:Accession: A55267
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <ZHE>
A:Cross-references: UNIPROT:Q28247; UNIPARC:UPI0000126D56; GB:U07888; NID:g469547; PIDN
C:Superfamily: collagen alpha 1(IV) chain

Query Match 81.0%; Score 34; DB 2; Length 754;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 189 GAGQPGGL 196

RESULT 46

S72526
inorganic diphosphatase (EC 3.6.1.1), H+-translocating, vacuolar membrane (clone OVP1)
C:Species: Oryza sativa (rice)
C>Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 27-Oct-2003
C:Accession: S72526
R;Sakakibara, Y.; Kobayashi, H.; Kasamo, K.
Plant Mol. Biol. 31, 1029-1038, 1996
A>Title: Isolation and characterization of cDNAs encoding vacuolar H(+)-pyrophosphatase
A:Reference number: S72526; MUID:97000915; PMID:8843945
A:Accession: S72526
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-771 <SAK>
A:Cross-references: UNIPARC:UPI00000A7481; EMBL:D45383; NID:g1747293; PIDN:BAA08232.1;
A:Note: only a part of the nucleic acid sequence is shown
C:Superfamily: H(+)-translocating inorganic pyrophosphatase
C:Keywords: hydrolase

Query Match 81.0%; Score 34; DB 2; Length 771;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
||| |||
DB 43 GEGGSPG 49

RESULT 47

T45467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45467
R;Richardson, D.W.; Dodge, G.R.
submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in e
A:Reference number: Z22977
A:Accession: T45467
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; UNIPARC:UPI000008834A; EMBL:U62528; PIDN:AAB05773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 81.0%; Score 34; DB 2; Length 1418;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
||| |||
DB 672 GAGQPGGL 679

```

RESULT 48
A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A44885
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:Cross-references: UNIPARC:UPI0000177381; GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P.
Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilag
A:Reference number: A44885; MUID:91347939; PMID:1879363
A:Accession: A44885
A:Molecule type: DNA
A:Residues: 1-28 <CHE>
A:Cross-references: UNIPARC:UPI00000004E7; GB:S63190; NID:9234368; PIDN:AAB19627.1; PID:
A:Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBIP:63192)
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      81.0%; Score 34; DB 2; Length 1419;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||
Db 673 GAQGPPGL 680

RESULT 49
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: B40333
R:Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A:Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A:Reference number: A40333; MUID:92011898; PMID:1918153
A:Accession: B40333
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1486 <SUA>
A:Cross-references: UNIPROT:Q91718; UNIPARC:UPI0000173B50; GB:M63595
C:Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F:37-96/Domain: von Willebrand factor type C repeat homology <WVC>
F:1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      81.0%; Score 34; DB 1; Length 1486;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||
Db 743 GAQGPPGL 750

RESULT 50
CGHUC6
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen

```

```

C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence_revision 01-Sep-1995 #text_change 31-Dec-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S63
7250; I37251; I37253; I37254; I55338; I55335; I61910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-cod
A:Reference number: A38513; MUID:91184811; PMID:2081599
A:Accession: A38513
A:Molecule type: DNA
A:Residues: 1-103 <RYA>
A:Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
:918084
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A:Reference number: S06715; MUID:90067946; PMID:2587267
A:Accession: S06715
A:Molecule type: mRNA
A:Residues: 1-28, 'R', '99-1487 <SU2>
A:Cross-references: UNIPARC:UPI00000126D15; EMBL:X16468; NID:G29515; PIDN:CAA34488.1; PID
A:Note: alternative splice form 1
R:Vikkula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A:Reference number: S24270; MUID:92344585; PMID:1637314
A:Accession: S24270
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-28 <VKS>
A:Cross-references: UNIPARC:UPI0000173B64; EMBL:X58709; GB:S40537; NID:G35659
A:Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A:Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A:Reference number: A24828; MUID:87031574; PMID:3021582
A:Accession: A24828
A:Molecule type: DNA
A:Residues: 1-8, 'T', '10-28 <NUN>
A:Cross-references: UNIPARC:UPI0000016A71A; GB:M25698; NID:G180872; PIDN:AAA52051.1; PID:
R:Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A:Reference number: S06496; MUID:90026318; PMID:2803268
A:Accession: S06496
A:Molecule type: mRNA
A:Residues: 7-28, 'R', '99-157, 'P', '159-440, 'G', '442-456, 'E', '458-640, 'A', '642-831, 'PA', '834, 'F'
A:Cross-references: UNIPARC:UPI0000173B65; EMBL:X16711; NID:G30040; PIDN:CAA34683.1; PID
A:Note: alternative splice form 1
R:Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A:Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A:Reference number: A35428; MUID:90285153; PMID:2355003
A:Accession: A35428
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-81, 'L', '83-103 <RYA2>
A:Cross-references: UNIPARC:UPI00000173B66
A:Note: alternative splice form 2; splicing appears to be under developmental regulation
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A:Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A:Reference number: A30147; MUID:89233138; PMID:2714801
A:Accession: A30147
A:Molecule type: DNA
A:Residues: 104-157, 'P', '159-236 <SUM>
A:Cross-references: UNIPARC:UPI0000173B67; GB:J03065; GB:M23660; GB:M25655; GB:M25656; G
R:Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A:Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A:Reference number: A94227; MUID:90370826; PMID:1975693
A:Accession: A33116
A:Molecule type: DNA

```


A;Residues: 171-172,'C',174-175 <ALA>
A;Cross-references: UNIPARC:UPI0000173B68
A;Note: mutant sequence from a family with primary generalized osteoarthritis
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189,'X',191-195,1224-1230,'X',1232-1236 <DIA>
A;Cross-references: UNIPARC:UPI0000173B69; UNIPARC:UPI0000173B6A
R;Franc, S.; Marzin, E.; Bouillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil can
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261;575-590;756-763,'X',765-779 <FRA>
A;Cross-references: UNIPARC:UPI0000173B6B; UNIPARC:UPI0000173B6C; UNIPARC:UPI0000173B6D
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G>51VS20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DBBJ
A;Molecule type: DNA
A;Residues: 440,'G',442-456,'E',458-480,'P',482-509 <TII1>
A;Cross-references: UNIPARC:UPI00006F3AF; EMBL:U15195; NID:G557053; PIDN:AAB60370.1; PI
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676,'A',678-783,'A',785-831,'PA',834,'P',836-1214 <RAM>
A;Cross-references: UNIPARC:UPI000016A700; EMBL:X13783; NID:G30037; PIDN:CAA32030.1; PID
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.
A;Reference number: S05000; MUID:89325561; PMID:2753125
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640,'A',642-785 <VIR2>
A;Cross-references: UNIPARC:UPI0000173B6E; EMBL:X16158; NID:G29951; PIDN:CAA34278.1; PID
A4282.1; PID:G1335022; PIDN:CAA34283.1; PID:G1335023; PIDN:CAA34284.1; PID:G1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA
A;Residues: 752-831,'PA',834,'F',836-1005,'K',1007-1036,'Q',1038-1052,'E',1054-1068,'T',
A;Cross-references: UNIPARC:UPI0000173B6F; GB:I00977; NID:G180812; PIDN:AAB33914.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIP:117273); parts of this sequence we
A;Note: this translation is not annotated and this publication is not cited in GenBank
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184,'GPSKDGANGIPGP',1185-1199 <TII2>
A;Cross-references: UNIPARC:UPI000011F72; EMBL:M37126; NID:G180808; PIDN:AAA52037.1; PI
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, P.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056,'N',1058-1068,'T',1070-1487 <CHE>
A;Cross-references: UNIPARC:UPI000016A6BB; GB:J00116; NID:G180395; PIDN:AAA51997.1; PID:

R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: UNIPARC:UPI000016A71B; EMBL:X06268; NID:G30096; PIDN:CAA29604.1; PFI
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Cross-references: UNIPARC:UPI0000173B70; UNIPARC:UPI0000173B71; UNIPARC:UPI0000173B72
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal pro
Nucleic Acids Res. 12, 1025-1038, 1984
R;Strom, C.M.; Upholt, W.B.
A;Title: Isolation and characterization of genomic clones corresponding to the human ty
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: UNIPARC:UPI000016A61A; EMBL:X00339; EMBL:X00298; NID:G394699; PIDN:
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE', <STR2>
A;Cross-references: UNIPARC:UPI000006EA4D; GB:K01785; NID:G30035; PIDN:CAA25082.1; PID:
R;Nunez, A.M.; Francmano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pr
gene.
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: UNIPARC:UPI0000173B73; GB:M12048; NID:G180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.D.; Sobel, M.E.; Tsipouras, P.; Ramirez
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) col
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28,'R',99-114;541-578;786-802;1055-1056,'N',1058-1068,'T',1070-1109;1200-
A;Cross-references: UNIPARC:UPI0000071780; UNIPARC:UPI0000173B74; UNIPARC:UPI0000173B75
A;Accession: I84453
A;Status: translated from GB/EMBL/DBBJ
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>
A;Cross-references: UNIPARC:UPI0000071780; UNIPARC:UPI0000173B74; UNIPARC:UPI0000173B75
24938; NID:G30104
A;Note: the GenBank PID is based on an incorrect reading frame
Query Match 81.0%; Score 34; DB 1; Length 1487;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
DB 741 GAQGPGL 748
RESULT 51
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 31-Dec-2004
C;Accession: B41182
R;Metaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 266, 16862-16869, 1991
 A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
 A;Reference number: A41182; MUID:91358489; PMID:1885613
 A;Accession: B41182
 A;Status: preliminary; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 1-1487 <MET>
 A;Cross-references: UNIPROT:Q62031; UNIPROT:Q62032; UNIPROT:Q62033; UNIPARC:UPI000017737
 C;Superfamily: fibrillar collagen carboxyl-terminal homology; von Willebrand factor type
 C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
 F;33-91/Domain: von Willebrand factor type C repeat homology <VMC>
 F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 81.0%; Score 34; DB 2; Length 1487;
 Best Local Similarity 75.0%; Pred. No. 3.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 ||:||||
 Db 741 GAQGPPL 748

RESULT 52
 A40333
 collagen alpha 1'(II) chain precursor - African clawed frog
 C;Species: Xenopus laevis (African clawed frog)
 C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
 C;Accession: A40333
 J;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
 J. Cell Biol. 115, 565-575, 1991
 A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
 A;Reference number: A40333; MUID:92011898; PMID:1918153
 A;Accession: A40333
 A;Status: nucleic acid sequence not shown
 A;Molecule type: mRNA
 A;Residues: 1-1492 <SUA>
 A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; UNIPARC:UPI0000177382; GB:M63596
 A;Note: this sequence is presented as substitutions relative to another sequence in a fi
 es they replace; the appropriate interpretation of the sequence figure was reconstructed
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
 F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
 F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 81.0%; Score 34; DB 2; Length 1492;
 Best Local Similarity 75.0%; Pred. No. 3.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 ||:||||
 Db 746 GAQGPPL 753

RESULT 53
 I48103
 type VII collagen - Chinese hamster (fragment)
 C;Species: Cricetulus griseus (Chinese hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
 C;Accession: I48103
 R;Greenpan, D.S.
 Hum. Mol. Genet. 2, 273-278, 1993
 A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
 A;Reference number: I48103; MUID:93271985; PMID:8499916
 A;Accession: I48103
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1549 <RES>
 A;Cross-references: UNIPROT:Q60444; UNIPARC:UPI00000E753D; GB:I06863; NID:G388624; PIDN:
 F;1484-1536/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match 81.0%; Score 34; DB 2; Length 1549;
 Best Local Similarity 75.0%; Pred. No. 4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 ||:||||
 Db 838 GPQGPGL 845

RESULT 54
 CGHU48
 collagen alpha 1(IV) chain precursor - human
 N;Alternate names: procollagen alpha 1(IV) chain
 C;Species: Homo sapiens (man)
 C;Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
 C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
 R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 264, 13565-13571, 1989
 A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
 A;Reference number: S16876; MUID:89340433; PMID:2701944
 A;Accession: S16876
 A;Status: nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-1669 <SO11>
 A;Cross-references: UNIPROT:P02462; UNIPARC:UPI000004981D; EMBL:J04217; GB:J05039; NID:G
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
 R;Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
 J. Biol. Chem. 263, 17217-17220, 1988
 A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
 A;Reference number: A92690; MUID:89034231; PMID:3182844
 A;Accession: A32117
 A;Molecule type: DNA
 A;Residues: 1-28 <SO12>
 A;Cross-references: UNIPARC:UPI0000173BC1; EMBL:J04217; NID:gi80759; PIDN:AAA53097.1; PI
 R;Poeschl, E.; Pollner, R.; Kuehn, K.
 EMBO J. 7, 2687-2695, 1988
 A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane
 A;Reference number: S02738; MUID:89030632; PMID:2846280
 A;Accession: S02738
 A;Status: translation not shown
 A;Molecule type: DNA
 A;Residues: 1-6, 'L', 8-28 <POE>
 A;Cross-references: UNIPARC:UPI0000173BC2; EMBL:X12784; NID:G30072
 R;Brazel, D.; Oberbaumer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
 Eur. J. Biochem. 168, 529-536, 1987
 A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
 A;Reference number: S00048; MUID:88029471; PMID:3311751
 A;Accession: S00048
 A;Molecule type: mRNA
 A;Residues: 1-318, 'A', 320-944 <BRAL>
 A;Cross-references: UNIPARC:UPI000016A708; EMBL:X05561; NID:G30066; PIDN:CAA29075.1; PID
 A;Accession: S25826
 A;Molecule type: protein
 A;Residues: 271-318, 'A', 320-554 <BRA2>
 A;Cross-references: UNIPARC:UPI0000173BC3
 R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
 Eur. J. Biochem. 152, 213-219, 1985
 A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S
 A;Reference number: A23115; MUID:86004708; PMID:4043082
 A;Accession: A23115
 A;Molecule type: protein
 A;Residues: 28-236, 'K', 239-240, 'K', 242-243 <GLA>
 A;Cross-references: UNIPARC:UPI0000173BC4
 A;Experimental source: placenta
 A;Note: the amino end of the mature form is blocked
 R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
 FEBS Lett. 225, 188-194, 1987
 A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty
 A;Reference number: S00207; MUID:88083584; PMID:3691802
 A;Accession: S00207
 A;Molecule type: mRNA
 A;Residues: 244-530 <SO13>
 A;Cross-references: UNIPARC:UPI0000173BC5; EMBL:Y00706; NID:G29548; PIDN:CAA68698.1; PI
 R;Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
 EMBO J. 12, 4795-4802, 1993
 A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen

A;Reference number: S39614; MUID:94038963; PMID:8223488
 A;Accession: S39614
 A;Molecule type: protein
 A;Residues: 371-554 <EBL>
 A;Cross-references: UNIPARC:UPI0000173BC6
 R;Babel, W.; Glanville, R.W.
 Eur. J. Biochem. 143, 545-556, 1984
 A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sequence
 A;Reference number: A02863; MUID:85003629; PMID:6434307
 A;Accession: A02863
 A;Molecule type: protein
 A;Residues: 534-718; 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-1000
 A;Cross-references: UNIPARC:UPI0000173BC7
 R;Experimental source: placenta
 R;Glanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A;Title: Peptin fragments of human placental basement-membrane-collagens showing interrupted
 A;Reference number: S16908; MUID:82005835; PMID:6792033
 A;Accession: A58517
 A;Molecule type: protein
 A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-1412
 A;Cross-references: UNIPARC:UPI0000173BC8; UNIPARC:UPI0000173BC9
 R;MacKnight, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A;Title: Isolation and characterization of pepsin-solubilized human basement membrane (B
 A;Reference number: S16910; MUID:84053346; PMID:6416291
 A;Accession: S16910
 A;Molecule type: protein
 A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'V', 946, 'X', 948-949
 A;Cross-references: UNIPARC:UPI0000173BCA; UNIPARC:UPI0000173BCB
 R;Experimental source: placenta
 R;Pihaajanien, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; H
 J. Biol. Chem. 260, 7681-7687, 1985
 A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen
 A;Reference number: S01466; MUID:85207819; PMID:2581969
 A;Accession: S01466
 A;Molecule type: mRNA
 A;Residues: 1256-1669 <PIH>
 A;Cross-references: UNIPARC:UPI000016A6B8; EMBL:M10940; NID:G180421; PIDN:AAA52006.1; PI
 R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.; P
 Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
 A;Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
 A;Reference number: S16879; MUID:85216555; PMID:2582422
 A;Accession: S16879
 A;Molecule type: mRNA
 A;Residues: 1259-1669 <BRI>
 A;Cross-references: UNIPARC:UPI000016A707; EMBL:M11315; NID:G180817; PIDN:AAA52042.1; PI
 R;Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
 Eur. J. Biochem. 147, 217-224, 1985
 A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
 A;Reference number: A02864; MUID:85127033; PMID:2578961
 A;Accession: S19091
 A;Molecule type: protein
 A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491; 1501-1514, 'X', 1516-1519; 1534-1553, 'X',
 A;Cross-references: UNIPARC:UPI0000173BCB; UNIPARC:UPI0000173BCD; UNIPARC:UPI0000173BCE;
 R;Stebold, B.; Deutzmann, R.; Kuehn, K.
 Eur. J. Biochem. 176, 617-624, 1988
 A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
 A;Reference number: S02550; MUID:89005112; PMID:2844531
 A;Contents: annotation; disulfide bonds
 C;Genetics:
 A;Gene: GDB:COL4A1
 A;Map position: 13q34-13q34
 A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 147/3; 156/3; 184/3; 205/3; 217/3; 231/
 1/1; 731/3; 782/1; 820/1; 876/1; 906/1; 957/1; 990/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
 C;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2(IV)
 oiations among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
 r-trimer associations in the interrupted helical domain (with disulfide and desmosine cr
 C;Function:
 A;Description: structural component of extracellular basement membrane
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication

F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>
 F;163-1440/Domain: interrupted helical <COL>
 F;414-452/Region: integrin binding #status experimental
 F;597-599/Region: cell attachment (R-G-D) motif
 F;917-919/Region: cell attachment (R-G-D) motif
 F;968-970/Region: cell attachment (R-G-D) motif
 F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
 F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
 F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
 F;127/Modified site: blocked amino end (Ala) (in mature form) #status experimental
 F;31.36.39.41.125.434.467.470/Disulfide bonds: interchain #status predicted
 F;45.48.78.90.129.156.172.217.228.231.277.295.298.322.343.361.460.463.497.527.540.543.5
 1081.1084.1099.1117.1132.1150.1165.1185.1188.1206.1235.1265.1283.1304.1319.1328.13
 F;45.48.78.90.129.156.217.228.231.277.295.298.322.343.361.460.463.497.527.540.543.5
 99.1117.1132.1150.1165.1182.1185.1188.1206.1235.1265.1283.1304.1319.1328.1340.1356.1371.
 F;54.63.75.84.87.96.102.105.108.111.117.120.123.138.141.147.150.153.159.167.178.181.184.
 ,419.422.425.439.445.448.451.479.485.491.494.503.512.518.524.530.546.549.552.555.561.56.
 9.745.748.751.754.763/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;126/Binding site: carboxylate (Asn) (covalent) #status experimental
 F;129/Modified site: alllysine (Lys) #status predicted
 F;172.540.947/Modified site: 5-hydroxylysine (Lys) #status atypical
 F;272.645.839/Modified site: 4-hydroxyproline (Pro) #status atypical
 F;446-447/Cleavage site: Gly-Ile (gelatinase B) #status predicted
 F;766.775.784.787.790.796.799.804.810.816.822.834.860.863.869.872.875.887.890.893.899.9
 23.1129.1138.1141.1159.1171.1176.1179.1194.1200.1203.1215.1224.1227.1244.1247.1250.1256.
 431.1437/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;1120.1468/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F;1120.1468/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental
 F;1214.1424/Modified site: 3-hydroxyproline (Pro) #status absent
 F;1392.1395.1398.1404/Modified site: 4-hydroxyproline (Pro) #status experimental
 F;1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F;1505-1511.1616-1622/Disulfide bonds: #status predicted
 F;1570-1662.1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 81.0%; Score 34; DB 1; Length 1669;
 Best Local Similarity 75.0%; Pred. No. 4.3e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 DB 1032 GPQSPGL 1039

RESULT 55
 CGMS4B
 collagen alpha 1(IV) chain precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 28-May-1986 #sequence revision 31-Dec-1992 #text change 09-Jul-2004
 C;Accession: A33525; S01454; A28066; A02864; A25636; A29301; S19079; A32003; A31766; S1
 R;Muthukumar, G.; Blumberg, B.; Kurkinen, M.
 J. Biol. Chem. 264, 6310-6317, 1989
 A;Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Dif
 A;Reference number: A33525; MUID:89197932; PMID:2703490
 A;Accession: A33525
 A;Molecule type: mRNA
 A;Residues: 1-1669 <MUT>
 A;Cross-references: UNIPROT:P02463; UNIPARC:UPI000002724F; EMBL:J04694; NID:G556296; PI
 R;Wood, L.; Theriault, N.; Vogeli, G.
 FEBS Lett. 227, 5-8, 1988
 A;Title: cDNA clones completing the nucleotide and derived amino acid sequence of the a
 A;Reference number: S01454; MUID:88112221; PMID:3338568
 A;Accession: S01454
 A;Molecule type: mRNA
 A;Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-4
 A;Cross-references: UNIPARC:UPI0000173BD1; EMBL:X06777
 R;Kilken, P.D.; Burelo, P.; Sakurai, Y.; Yamada, Y.
 J. Biol. Chem. 263, 8706-8709, 1988
 A;Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha
 A;Reference number: A28066; MUID:88243724; PMID:3379041
 A;Accession: A28066
 A;Molecule type: mRNA

A;Residues: 1-129 <K1>
A;Cross-references: UNIPARC:UPI000016CCB0; EMBL:J03758; NID:g192669; PIDN:AAA37439.1; PI
R;Oberbauer, I.; Laurent, M.; Schwaerz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, E.
Eur. J. Biochem. 147, 217-224, 1985
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A;Reference number: A02864; MUID:85127033; PMID:2578961
A;Accession: A02864
A;Molecule type: mRNA
A;Residues: 1276-1669 <OB>
A;Cross-references: UNIPARC:UPI000016CC4B; EMBL:X02201; NID:g50233; PIDN:CAA26132.1; PID
R;Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
Gene 43, 301-304, 1986
A;Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
A;Reference number: A25636; MUID:86301886; PMID:3755692
A;Accession: A25636
A;Molecule type: mRNA
A;Residues: 1149-1396, 'S', 1398-1424 <NAT>
A;Cross-references: UNIPARC:UPI000016CC50; EMBL:M14042; NID:g192286; PIDN:AAA37342.1; PI
A;Note: the authors translated the codon CAG for residue 1374 as Arg
R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj
J. Biol. Chem. 262, 8496-8499, 1987
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
A;Reference number: A34680; MUID:87250460; PMID:3597383
A;Accession: A29301
A;Molecule type: mRNA
A;Residues: 1441-1669 <KUR>
A;Cross-references: UNIPARC:UPI000016CC4A; EMBL:M15832; NID:g192282; PIDN:AAA37340.1; PI
R;Killen, P.D.; Burbello, P.D.; Martin, G.R.; Yamada, Y.
J. Biol. Chem. 263, 12310-12314, 1988
A;Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
A;Reference number: S19079; MUID:86315019; PMID:2842328
A;Accession: S19079
A;Molecule type: DNA
A;Residues: 1-28 <K12>
A;Cross-references: UNIPARC:UPI000000038F; EMBL:J03944; NID:g192673; PIDN:AAA37442.1; PI
R;Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.
J. Biol. Chem. 263, 19274-19277, 1988
A;Title: Head-to-head arrangement of murine type IV collagen genes.
A;Reference number: A92702; MUID:89066738; PMID:3198626
A;Accession: A32003
A;Molecule type: DNA
A;Residues: 1-28 <KAY>
A;Cross-references: UNIPARC:UPI000000038F; EMBL:J04448; NID:g192666; PIDN:AAA37437.1; PI
R;Burbello, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promo
A;Reference number: A94220; MUID:89071759; PMID:3200851
A;Accession: A31766
A;Molecule type: DNA
A;Residues: 1-28 <BUR>
A;Cross-references: UNIPARC:UPI000000038F; EMBL:M23333; NID:g340878; PIDN:AAA51625.1; PI
R;Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
A;Reference number: S19094; MUID:86196099; PMID:3009468
A;Accession: S19094
A;Molecule type: DNA
A;Residues: 1110-1135; 1189-1316; 1342-1383; 1418-1487 <SAK>
A;Cross-references: UNIPARC:UPI000016CC51; UNIPARC:UPI000016CC52; UNIPARC:UPI000016CC53;
R;Schuppan, D.; Timpl, R.; Glangville, R.W.
FEBS Lett. 115, 297-300, 1980
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
A;Reference number: S16909; MUID:80246483; PMID:6772473
A;Accession: S16909
A;Molecule type: protein
A;Residues: 940-946, 'G', 951-955, 'G', 957-1213-1228, 'X', 1230-1234, 'P', 1236-123
A;Cross-references: UNIPARC:UPI0000173BD2; UNIPARC:UPI0000173BD3; UNIPARC:UPI0000173BD4
R;Schuppan, D.; Glangville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
A;Reference number: A25991; MUID:82186723; PMID:6804236
A;Accession: A25991
A;Molecule type: protein

A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 100
61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119,
A;Cross-references: UNIPARC:UPI0000173BD5
A;Accession: B25991
A;Molecule type: protein
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234,
3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1329,
A;Cross-references: UNIPARC:UPI0000173BD5
R;Weber, S.; Engel, J.; Wiedemann, H.; Glangville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A;Title: Subunit structure and assembly of the globular domain of basement-membrane coll
A;Reference number: S17801; MUID:84132058; PMID:6698021
A;Accession: S17801
A;Molecule type: protein
A;Residues: 1435-1443 <WEB>
A;Cross-references: UNIPARC:UPI0000173BDE
C;Genetics:
A;Note: the list of introns may be incomplete
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat
F;1-27/Domain: signal sequence #status predicted <SIG>
F;128-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;128-162/Domain: 7S <7SD>
F;163-1440/Domain: collagenous, triple helix <COL>
F;597-599/Region: cell attachment (R-G-D) motif
F;781-783/Region: cell attachment (R-G-D) motif
F;917-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1441-1552/Region: duplication
F;1553-1669/Region: duplication
F;31.36.39.41.434.467.470/Disulfide bonds: interchain #status predicted
F;128/Binding site: carboxydrate (Asn) (covalent) #status predicted
F;7971.974.977.986.989.1001.1007.1019.1022.1031.1037.1040.1055.1060.1063.1075.1078.1090.1
92.1298.1310.1313.1322.1337.1346.1349.1422.1425.1431.1437.1440/Modified site: hydroxypro
F;1214.1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1505-1511.1616-1622/Disulfide bonds: #status predicted
Query Match 81.0%; Score 34; DB 1; Length 1669;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 CAEGSPGL 8
| : |||||
Db 513 GPQSPGL 520
RESULT 56
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N;Alternate names: procollagen alpha 5(IV) chain
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 09-Jul-2004
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
R;Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and ident
n Allport syndrome patient.
A;Reference number: S22917; MUID:92316923; PMID:1352287
A;Accession: S22917
A;Molecule type: mRNA
A;Residues: 1-967 <ZHO>
A;Cross-references: UNIPROT:P29400; UNIPARC:UPI0000173BDF; GB:M90464; NID:g180826; PIDN:
R;Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A;Title: Structure of the human type IV collagen COL4A5 gene.
A;Reference number: A54365; MUID:94165049; PMID:8120014
A;Accession: A54365
A;Molecule type: DNA
A;Residues: 1-922 <ZH2>

A;Cross-references: UNIPARC:UPI0000173BE0; GB:U04470; NID:G463378; GB:U04520; NID:G46342
R;Zhou, J.; Mochizuki, T.; Sneets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggvason, S.
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alphas5(IV) and alpha6(IV) collagen genes in inherited sm
A;Reference number: A57079; MUID:93361972; PMID:8356449
A;Accession: A57079
A;Molecule type: DNA
A;Residues: 1-27 <ZK4>
A;Cross-references: UNIPARC:UPI000007378A; GB:Z37153; NID:G587203; PIDN:CAA85512.1; PID:
R;Pihlajaniemi, T.; Pohjolainen, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A;Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A;Reference number: A37122; MUID:90337990; PMID:2380186
A;Accession: A37122
A;Molecule type: mRNA
A;Residues: 84-439, 'GS', 442-624, 'LALO', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
A;Cross-references: UNIPARC:UPI0000173BE1; GB:J05558; EMBL:M58526; NID:G1314209
A;Note: submitted to the EMBL Data Library, February 1991
A;Note: the authors translated the codon GCC for residue 115 as Val
R;Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
Hum. Mol. Genet. 1, 127-129, 1992
A;Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
A;Reference number: 154317; MUID:93244772; PMID:1363780
A;Accession: 154317
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 313-324, 'E', 326-330 <REN>
A;Cross-references: UNIPARC:UPI000016B3D0; GB:SS9334; NID:G299946; PIDN:AAD13909.1; PID:
R;Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A;Title: Identification of a distinct type IV collagen alpha chain with restricted kidne
A;Reference number: A34850; MUID:90160375; PMID:1689491
A;Accession: A34850
A;Molecule type: mRNA
A;Residues: 914-1264, 1271-1691 <HOS>
A;Cross-references: UNIPARC:UPI000016A70B; EMBL:M31115; NID:G180824; PIDN:AAAS2045.1; PI
R;Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
Genomics 9, 1-9, 1991
A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A;Reference number: A37969; MUID:91169491; PMID:2004755
A;Accession: S18850
A;Molecule type: DNA
A;Residues: 924-1264, 1271-1691 <ZH3>
A;Cross-references: UNIPARC:UPI000016A437; EMBL:M63455; EMBL:M63457; EMBL:M
EMBL:M63467; EMBL:M63468; EMBL:M63470; EMBL:M63471; EMBL:M63473; NID:G17792
R;Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; V
Kidney Int. 44, 1316-1321, 1993
A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
A;Reference number: 156971; MUID:94133540; PMID:8301933
A;Accession: 156971
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1258-1276 <GUO1>
A;Cross-references: UNIPARC:UPI000016B421; GB:S69168; NID:G545095; PIDN:AACG0612.1; PID:
A;Note: kidney splice form
A;Accession: 176598
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1284-1291, 'TFLGYLACLV' <GUO2>
A;Cross-references: UNIPARC:UPI000011DDPD; GB:S69169; NID:G545097; PIDN:AACG0613.1; PID:
A;Note: frameshift mutation in patient with Alport syndrome
R;Myers, J.C.; Jones, T.A.; Pohjolainen, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; S
Am. J. Hum. Genet. 46, 1024-1033, 1990
A;Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the regi
A;Reference number: A35335; MUID:90252791; PMID:2339699
A;Accession: A35335
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1448-1477 <MYE>
A;Cross-references: UNIPARC:UPI0000173BE2
R;Nakatani, H.; Hattori, S.; Ushijima, T.; Matsaura, T.; Koitabashi, Y.; Takada, T.; Yos
Kidney Int. 46, 1307-1314, 1994
A;Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord

A;Reference number: 156975; MUID:95156893; PMID:7853788
A;Accession: 156975
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1595-1602 <NAK>
A;Cross-references: UNIPARC:UPI00000004F8; GB:S75903; NID:G913882; PIDN:AAB33374.1; PID:
R;Lemink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.;
Genomics 17, 485-489, 1993
A;Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
A;Reference number: 154188; MUID:94010948; PMID:8406498
A;Accession: 154188
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1604-1607, 'VHDAYKC' <LEM>
A;Cross-references: UNIPARC:UPI000011F85C; GB:S65767; NID:G425563; PIDN:AAD13967.1; PID:
A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations ar
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A5; ATS
A;Cross-references: GDB:120596; OMIM:303630
A;Map position: Xq22-Xq22
A;Intons: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5(
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1
F;27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status F
F;27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F;42-1462/Region: interrupted helical
F;1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1583-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;29-32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F;125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F;1527-1533,1638-1644/Disulfide bonds: #status predicted
F;1592-1684,1626-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted
Query Match 81.0%; Score 34; DB 1; Length 1691;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAGGSPGL 8
DB 1113 GAGGQFGL 1120
RESULT 57
A54121
collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
N;Alternate names: collagen alpha 2(IV) chain homolog
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004
C;Accession: A54121; S44317
R;Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Ramirez, F.
J. Biol. Chem. 269, 13167-13171, 1994
A;Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I
A;Reference number: A54121; MUID:94230414; PMID:8175744
A;Accession: A54121
A;Molecule type: mRNA
A;Residues: 1-1747 <EXP>
A;Cross-references: UNIPROT:Q26640; UNIPARC:UPI0000076901; EMBL:X76730; NID:G483606; PI
C;Genetics:
A;Gene: COLP4alpha

C:Superfamily: collagen alpha 1(IV) chain

Query Match 81.0%; Score 34; DB 2; Length 1747;
 Best Local Similarity 75.0%; Pred. No. 4.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 591 GPDGSPGL 598

RESULT 58

S16366 collagen alpha 2(IV) chain precursor - pig roundworm

C:Species: Ascaris suum (pig roundworm)
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
 C:Accession: S16366

R:Pettitt, J.; Kingston, I.B.
 J. Biol. Chem. 266, 16149-16156, 1991
 A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the partial complementary DNA sequence
 A:Reference number: S16366; MUID:91340768; PMID:1714907

A:Accession: S16366

A:Molecule type: mRNA

A:Residues: 1-1763 <JBI>

A:CROSS-references: UNIPROT:P27393; UNIPARC:UPI0000126D40; GB:M67507; NID:g159648; PIDN:

C:Genetics:

A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfide

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>

F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>

F:43-1529/Domain: collagenous #status predicted <COL>

F:1530-1763/Domain: carboxyl-terminal nonhelical, NCl #status predicted <NCl1>

F:1530-1638/Domain: repeat NCl #status predicted <NCl1>

F:1639-1763/Domain: repeat NCl #status predicted <NCl2>

F:31.34,39,41,536,539/Disulfide bonds: interchain #status predicted

F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 81.0%; Score 34; DB 2; Length 1763;

Best Local Similarity 75.0%; Pred. No. 4.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 879 GAKGEPGL 886

RESULT 59

A31893

C:Superfamily: collagen alpha 1(IV) chain precursor - fruit fly (Drosophila melanogaster)

C:Species: Drosophila melanogaster

C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 09-Jul-2004

C:Accession: A31893; A26692; A19442; S00020

R:Blumberg, B.; MacKrell, A.J.; Fessler, J.H.

J. Biol. Chem. 263, 18328-18337, 1988

A:Title: Drosophila basement membrane procollagen alpha-1(IV). II. Complete cDNA sequence

A:Reference number: A31893; MUID:89054012; PMID:3142875

A:Accession: A31893

A:Molecule type: mRNA

A:Residues: 1-1775 <BLU>

A:CROSS-references: UNIPROT:P08120; UNIPARC:UPI0000126D1C; EMBL:M23704; NID:g157029; PID

R:Blumberg, B.; MacKrell, A.J.; Olson, P.F.; Kurkinen, M.; Monson, J.M.; Natzle, J.E.; F

J. Biol. Chem. 262, 5947-5950, 1987

A:Title: Basement membrane procollagen IV and its specialized carboxyl domain are conserved

A:Reference number: A26692; MUID:87194801; PMID:3106346

A:Accession: A26692

A:Molecule type: mRNA

A:Residues: 1065-1775 <BLU2>

A:CROSS-references: UNIPARC:UPI00001773AF; EMBL:J02727

R:Monson, J.M.; Natzle, J.; Friedman, J.; McCarthy, B.J.

Proc. Natl. Acad. Sci. U.S.A. 79, 1761-1765, 1982
 A:Title: Expression and novel structure of a collagen gene in Drosophila.
 A:Reference number: A19442; MUID:82197577; PMID:6210912

A:Accession: A19442

A:Molecule type: DNA

A:Residues: 762-947, 'S', 949-996, 'T', 998-1230 <MON>

A:CROSS-references: UNIPARC:UPI000016BB4B; GB:J01074; EMBL:V00200; NID:g7736; PIDN:CAA23A

R:Cecchini, J.P.; Knibiehler, B.; Mirre, C.; le Parco, Y.

Eur. J. Biochem. 165, 587-593, 1987

A:Title: Evidence for a type-IV-related collagen in Drosophila melanogaster. Evolutionary

A:Reference number: S00020; MUID:87246644; PMID:3109906

A:Accession: S00020

A:Molecule type: DNA

A:Residues: 1355-1356, 'K', 1358-1359, 'K', 1361-1372, 'I', 1374-1495, 'R', 1497-1506, 'RA', 1509,

A:CROSS-references: UNIPARC:UPI00001773B0; EMBL:M28334

C:Genetics:

A:Gene: FlyBase:Cg25C

A:CROSS-references: FlyBase:FBgn0000299

A:Introns: 7/2; 23/3; 339/3; 505/2; 989/1; 1312/1; 1685/3

C:Superfamily: collagen alpha 1(IV) chain

C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; ex

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-1775/Product: collagen alpha 1(IV) chain #status predicted <MAT>

F:65-67/Region: cell attachment (R-G-D) motif

F:130-132/Region: cell attachment (R-G-D) motif

F:238-240/Region: cell attachment (R-G-D) motif

F:297-299/Region: cell attachment (R-G-D) motif

F:892-894/Region: cell attachment (R-G-D) motif

F:1075-1077/Region: cell attachment (R-G-D) motif

F:1173-1175/Region: cell attachment (R-G-D) motif

F:1225-1227/Region: cell attachment (R-G-D) motif

F:1545-1545/Domain: carboxyl-terminal nonhelical, NCl #status predicted <NCl1>

F:1545-1655/Domain: repeat NCl #status predicted <NCl1>

F:1656-1775/Domain: repeat NCl #status predicted <NCl2>

F:72/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:570,573/Disulfide bonds: interchain #status predicted

F:1611-1617,1720-1727/Disulfide bonds: #status predicted

Query Match 81.0%; Score 34; DB 2; Length 1775;

Best Local Similarity 75.0%; Pred. No. 4.6e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 672 GAGLPGI 679

RESULT 60

A54849

C:Superfamily: collagen alpha 1(VII) chain precursor - human

C:Species: Homo sapiens (man)

C:Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 09-Jul-2004

C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686

R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.

J. Biol. Chem. 269, 20256-20262, 1994

A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII

A:Reference number: A54849; MUID:94327588; PMID:8051117

A:Accession: A54849

A:Molecule type: mRNA

A:Residues: 1-2944 <CHR>

A:CROSS-references: UNIPROT:Q02388; UNIPARC:UPI000017A138; GB:L02870; NID:g987124; PIDN:

R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.

Biochem. Biophys. Res. Commun. 183, 958-963, 1992

A:Title: Molecular cloning and characterization of type VII collagen cDNA.

A:Reference number: PH0844; MUID:92231902; PMID:1567409

A:Accession: PH0844

A:Molecule type: mRNA

A:Residues: 'EFR', 340-475, 'RALSTASHSTLCWTRATRHPCNRGSHWTFRAACEPCNRPASHRAARAG', 524-528, 'C',

A:CROSS-references: UNIPARC:UPI000017A139; DBJ:DJ13694; NID:g453698; PIDN:BJ

A:Experimental source: keratinocyte

A>Note: the authors translated the codon ACC for residues 394 and 397 as Tyr

R;Parente, M.G.; Chung, L.C.; Ryyanaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A;Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A;Reference number: S16316; MUID:91334380; PMID:1871109
A;Accession: S16316
A;Molecule type: mRNA
A;Residues: 815-892, 'E', 894-1439 <PAR>
A;Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:G180914; PIDN:AAA96
A;Experimental source: keratinocyte
R;Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Priayanh, P.S.; Cook, M.E.; Wright, J.;
J. Invest. Dermatol. 99, 691-696, 1992
A;Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proc
A;Reference number: I56328; MUID:91107742; PMID:1469284
A;Accession: I56328
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 'ER', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>
A;Cross-references: UNIPARC:UPI000016B3AC; GB:S51236; NID:G262308; PIDN:AAB24637.1; PID:
R;Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Ghanville, R.W.; Burgeson, R.E.
J. Biol. Chem. 264, 3822-3826, 1989
A;Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagen
A;Reference number: A30296; MUID:89139437; PMID:2537292
A;Accession: A30296
A;Molecule type: protein
A;Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041;
A;Cross-references: UNIPARC:UPI000017A13A; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
A;Note: two reported peptides cannot be reliably located
R;Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A;Reference number: I48103; MUID:93271985; PMID:8499916
A;Accession: I48103
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 2395-2871, 'S', 2873-2944 <RE2>
A;Cross-references: UNIPARC:UPI000016A724; GB:L06862; NID:G388713; PIDN:AAA89196.1; PID:
R;Christiano, A.M.; Ryyanaenen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A;Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser sube
A;Reference number: A55255; MUID:94224777; PMID:8170945
A;Comments: annotation
C;Content: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL7A1; EBR1; EBD1; EB
A;Cross-references: GDB:128750; OMIM:120120
A;Map position: 3p21.3-3p21.3
A;Note: defects in this Gene can result in dominant and recessive dystrophic epidermolys
C;Complex: type VII collagen is probably a homotrimer
C;Function:
A;Description: structural component of extracellular polymer associated with anchoring f
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F;17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F;17-201/Domain: von Willebrand factor type A repeat homology <VW1>
F;231-318/Domain: fibronectin type III repeat homology <FN1>
F;327-413/Domain: fibronectin type III repeat homology <FN2>
F;414-502/Domain: fibronectin type III repeat homology <FN3>
F;508-593/Domain: fibronectin type III repeat homology <FN4>
F;598-683/Domain: fibronectin type III repeat homology <FN5>
F;686-771/Domain: fibronectin type III repeat homology <FN6>
F;776-862/Domain: fibronectin type III repeat homology <FN7>
F;864-952/Domain: fibronectin type III repeat homology <FN8>
F;954-1045/Domain: fibronectin type III repeat homology <FN9>
F;1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>
F;1170-1172/Region: cell attachment (R-G-D) motif
F;1189-1253/Region: cysteine/proline-rich
F;1254-2783/Region: interrupted helical
F;1334-1336/Region: cell attachment (R-G-D) motif
F;2008-2010/Region: cell attachment (R-G-D) motif
F;2553-2555/Region: cell attachment (R-G-D) motif

F;2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F;2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F;337,786,1109/Binding site: carboxylate (Asn) (covalent) #status predicted
F;2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F;2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F;2625,2631/Binding site: carboxylate (Lys) (covalent) #status experimental
F;2634,2802,2804/Diulfide bonds: interchain #status predicted

Query Match 81.0%; Score 34; DB 2; Length 2944;

Best Local Similarity 75.0%; Pred. No. 7.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8

|:|||||

DB 2239 GPQGSPL 2246

RESULT 61

B83173

hypothetical protein PA3786 [imported] - Pseudomonas aeruginosa (strain PAO1)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004

C;Accession: B83173

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Lam,
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho

A;Reference number: A82950; MUID:20437337; PMID:10984043

A;Accession: B83173

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-129 <STO>

A;Cross-references: UNIPROT:O9HXK9; UNIPARC:UPI00000C5B11; GB:AE004797; GB:AE004091; NIT

A;Experimental source: strain PAO1

C;Genetics:

A;Gene: PA3786

Query Match 78.6%; Score 33; DB 2; Length 129;

Best Local Similarity 75.0%; Pred. No. 49;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8

|:|||||

DB 80 GAAGTPEGL 87

RESULT 62

T24064

hypothetical protein R09A8.4 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T24064

submitted to the EMBL Data Library, November 1995

A;Reference number: Z19836

A;Accession: T24064

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-152 <WIL>

A;Cross-references: UNIPROT:Q21855; UNIPARC:UPI000007928C; EMBL:Z68009; PIDN:CAA92006.1

A;Experimental source: clone R09A8

C;Genetics:

A;Gene: CESP:R09A8.4

A;Map position: X

Query Match 78.6%; Score 33; DB 2; Length 152;

Best Local Similarity 85.7%; Pred. No. 58;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPG 7

|:|||||

DB 37 GREGSPG 43

```

A;Cross-references: UNIPROT:O44173; UNIPARC:UPI000007983C; GB:chr_IV; PIDN:AA888359.1; P
A;Note: Similar to cuticular collagen
C;Genetics:
A;Gene: F58F6.2
A;Map position: 4

Query Match      78.6%; Score 33; DB 2; Length 290;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||
Db 255 GARGOPGL 262

RESULT 66
T18637
hypothetical protein B0024.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18637
R;McMurray, A.
submitted to the EMBL Data Library, April 1996
A;Reference number: Z19001
A;Accession: T18637
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-297 <WIL>
A;Cross-references: UNIPROT:Q17417; UNIPARC:UPI000007CC8B; EMBL:Z71178; PIDN:CAA94874.1;
A;Experimental source: clone B0024
C;Genetics:
A;Gene: CESP:B0024.1
A;Map position: 5
A;Introns: 39/3

Query Match      78.6%; Score 33; DB 2; Length 297;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||
Db 269 GARGNPGL 276

RESULT 67
I50696
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I50696
R;Nah, H.D.; Niu, Z.; Adams, S.L., 1994
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not enc
A;Reference number: A54041; MUID:94266842; PMID:8206952
A;Accession: I50696
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-310 <NAH>
A;Cross-references: UNIPROT:Q90612; UNIPARC:UPI00000FC2F0; EMBL:U07974; NID:9520456; PID
C;Genetics:
A;Gene: COL3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match      78.6%; Score 33; DB 2; Length 310;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
    |||
Db 155 GASGSPG 161

RESULT 68

```

```

hypothetical protein W03G11.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T26125
R;McMurray, A.
submitted to the EMBL Data Library, November 1995
A;Reference number: Z20156
A;Accession: T26125
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-279 <WIL>
A;Cross-references: UNIPROT:Q23150; UNIPARC:UPI000017A123; EMBL:Z67738; PIDN:CAA91544.1;
A;Experimental source: clone W03G11
C;Genetics:
A;Gene: CESP:W03G11.1
A;Map position: X

Query Match      78.6%; Score 33; DB 2; Length 279;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
    |||
Db 168 GARGSPG 174

RESULT 64
A32249
collagen - sea urchin (Paracentrotus lividus) (fragment)
C;Species: Paracentrotus lividus (common urchin)
C;Date: 17-Aug-1989 #sequence_revision 17-Aug-1989 #text_change 09-Jul-2004
C;Accession: A32249
R;Saitta, B.; Butrice, G.; Gambino, R.
Biochem. Biophys. Res. Commun. 158, 633-639, 1989
A;Title: Isolation of a putative collagen-like gene from the sea urchin Paracentrotus li
A;Reference number: A32249; MUID:89149773; PMID:2537631
A;Accession: A32249
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-290 <SAI>
A;Cross-references: UNIPROT:Q26054; UNIPARC:UPI0000177399
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match      78.6%; Score 33; DB 2; Length 290;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||
Db 247 GARGAPGL 254

RESULT 65
B88638
protein F58F6.2 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C;Accession: B88638
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A;Reference number: A75000; MUID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A;Accession: B88638
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-290 <STO>

```


C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004

C;Accession: S11449

R;Exposito, J.Y.; Ouazana, R.; Garrone, R.

Eur. J. Biochem. 190, 401-406, 1990

A;Title: Cloning and sequencing of a Porifera partial cDNA coding for a short-chain collagen

A;Reference number: S11449; MUID:90306040; PMID:2163843

A;Accession: S11449

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-366 <EXP>

A;Cross-references: UNIPROT:P18503; UNIPARC:UPI0000126FD6; EMBL:X52598; NID:g9297; PIDN:

Query Match 78.6%; Score 33; DB 2; Length 366;

Best Local Similarity 75.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 52 GAPGAPGL 59

RESULT 74

F96788

protein T4012.22 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C;Accession: F96788

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: F96788

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-369 <STO>

A;Cross-references: UNIPROT:Q9LQR7; UNIPARC:UPI00000A1605; GB:AE005173; NID:g8778814; PI

C;Genetics:

A;Gene: T4012.22

A;Map position: 1

Query Match

Best Local Similarity 85.7%; Score 33; DB 2; Length 369;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 162 GREGSPG 168

RESULT 75

T27806

hypothetical protein ZK265.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T27806

R;Dobson, R.

submitted to the EMBL Data Library, October 1996

A;Reference number: Z20422

A;Accession: T27806

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-381 <WIL>

A;Cross-references: UNIPROT:Q94399; UNIPARC:UPI000007BB4F; EMBL:Z81143; PIDN:CAB03513.1;

A;Experimental source: clone ZK265

C;Genetics:

A;Gene: ZK265.2

A;Map position: 1

A;Introns: 23/1

C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 78.6%; Score 33; DB 2; Length 381;

Best Local Similarity 75.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 191 GDEGTPGL 198

Search completed: March 11, 2006, 12:06:19

Job time : 29.8 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 11:52:06 ; Search time 117.2 Seconds
(without alignments)

48.159 Million cell updates/sec

Title: US-10-698-121A-1

Perfect score: 42

Sequence: 1 GAEGSPGL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database :

Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	594	2	Q99228 HUMAN
2	42	100.0	623	2	Q14035 HUMAN
3	42	100.0	652	2	Q5TAT5 HUMAN
4	42	100.0	683	2	Q5TAT4 HUMAN
5	42	100.0	717	2	Q9NQ52 HUMAN
6	42	100.0	717	2	Q5TAT6 HUMAN
7	39	92.9	405	2	Q4RT24 TETNG
8	39	92.9	856	2	Q4SB07 TETNG
9	39	92.9	1026	2	Q4SZ73 TETNG
10	39	92.9	1707	1	C04A2 MOUSE
11	38	90.5	164	2	Q5BW21 SCHJA
12	38	90.5	228	2	Q810M6 MOUSE
13	38	90.5	230	2	Q9R149 CAVPO
14	38	90.5	460	2	Q5R949 PONGY
15	38	90.5	589	2	Q99LL6 MOUSE
16	38	90.5	698	2	Q86728 BACRY
17	38	90.5	779	1	C01A1 BOVIN
18	38	90.5	887	1	ZFY28 HUMAN
19	38	90.5	1069	2	Q6LAN8 HUMAN
20	38	90.5	1433	1	C01A1 MOUSE
21	38	90.5	1453	2	Q63079 RAT
22	38	90.5	1453	2	Q810J9 MOUSE
23	38	90.5	1460	1	C01A1 CANFA
24	38	90.5	1461	2	Q76045 HUMAN
25	38	90.5	1464	1	C01A1 HUMAN
26	38	90.5	1464	2	Q8N473 HUMAN
27	38	90.5	1467	2	Q59F64 HUMAN
28	38	90.5	1688	2	Q597P9 CANFA
29	38	90.5	2553	2	Q7VWJ2 BORPE
30	37	88.1	435	2	Q5AQ8 EMENT
31	37	88.1	812	2	Q06452 EPHMU

32	88.1	998	2	Q8CFM4 MOUSE	Q8cfm4 mus musculus
33	88.1	1027	1	CAFF_RIFPA	P30754 riftia pach
34	88.1	1222	2	Q8K173 MOUSE	Q8k173 mus musculus
35	88.1	1324	2	Q53RW9 HUMAN	Q53rw9 homo sapien
36	88.1	1414	2	Q26634 STRPU	Q26634 strongyloce
37	88.1	1464	1	C03A1 MOUSE	P08121 mus musculus
38	88.1	1464	2	Q8BK12 MOUSE	Q8bk12 mus musculus
39	88.1	1464	2	Q8BLW4 MOUSE	Q8blw4 mus musculus
40	88.1	1464	2	Q7TT32 MOUSE	Q7tt32 mus musculus
41	88.1	1467	2	Q5DTG2 MOUSE	Q5dtg2 mus musculus
42	88.1	1690	1	C04A4 HUMAN	P53420 homo sapien
43	88.1	1690	2	Q5WRI HUMAN	Q53wr1 homo sapien
44	88.1	1726	2	Q4RWM6 TETNG	Q4rwm6 tetraodon n
45	85.7	95	2	Q5NC68 MOUSE	Q5nc68 mus musculus
46	85.7	150	2	Q9CZS2 MOUSE	Q9czs2 mus musculus
47	85.7	187	2	Q61246 MOUSE	Q61246 mus musculus
48	85.7	215	2	Q4NK30 9MICC	Q4nk30 arthrobacte
49	85.7	218	2	Q71RG9 HUMAN	Q71rg9 homo sapien
50	85.7	287	2	Q6A622 PROAC	Q6a622 propionibac
51	85.7	321	2	Q61TB7 CAEBR	Q61tb7 caenorhabdi
52	85.7	340	2	Q82LA4 STRAW	Q82la4 streptomyce
53	85.7	374	2	Q7G6D9 ORYSA	Q7g6d9 oryza sativ
54	85.7	374	2	Q8SB75 ORYSA	Q8sb75 oryza sativ
55	85.7	415	2	Q568Y4 RAT	Q568y4 rattus norv
56	85.7	447	2	Q16593 HUMAN	Q16593 homo sapien
57	85.7	486	1	EGR4 HUMAN	Q05215 homo sapien
58	85.7	531	2	Q5TKQ7 ORYSA	Q5tkq7 oryza sativ
59	85.7	565	2	Q8K036 MOUSE	Q8k036 mus musculus
60	85.7	568	2	Q8CD80 MOUSE	Q8cd80 mus musculus
61	85.7	575	2	Q8CIF9 MOUSE	Q8cif9 mus musculus
62	85.7	598	2	Q86TR4 HUMAN	Q86tr4 homo sapien
63	85.7	619	2	Q68220 ANAPH	Q68220 anaplasma p
64	85.7	650	2	Q17866 CAEEL	Q17866 caenorhabdi
65	85.7	678	2	Q93486 ONCMY	Q93486 oncorhynch
66	85.7	739	2	Q70575 MOUSE	Q70575 mus musculus
67	85.7	749	2	Q79F74 MYCTU	Q79f74 mycobacteri
68	85.7	749	2	Q79FV7 MYCTU	Q79fv7 mycobacteri
69	85.7	751	2	Q9R1N9 MOUSE	Q9rin9 mus musculus
70	85.7	773	2	Q7U160 MYCBO	Q7u160 mycobacteri
71	85.7	778	2	Q9U9K6 CAEEL	Q9u9k6 caenorhabdi
72	85.7	911	1	C0BA1 BOVIN	Q28083 bos taurus
73	85.7	915	2	Q6ZQK3 MOUSE	Q6zqk3 mus musculus
74	85.7	925	2	Q4SIU4 TETNG	Q4siu4 tetraodon n
75	85.7	983	2	Q501R9 RAT	Q501r9 rattus norv
76	85.7	988	1	NBR1 MOUSE	P97432 mus musculus
77	85.7	1017	2	Q59H85 HUMAN	Q59hb5 homo sapien
78	85.7	1117	2	Q9U9K7 CAEEL	Q9u9k7 caenorhabdi
79	85.7	1134	2	Q4SWH0 TETNG	Q4swh0 tetraodon n
80	85.7	1138	2	Q7JL30 CAEEL	Q7jl30 caenorhabdi
81	85.7	1142	2	Q6P3P9 XENTR	Q6p3p9 xenopus tro
82	85.7	1169	2	Q8KPW2 9ACTO	Q8kpw2 streptomyce
83	85.7	1175	2	Q810G6 CAEEL	Q810g6 caenorhabdi
84	85.7	1310	2	Q9H7L6 HUMAN	Q9h7l6 homo sapien
85	85.7	1352	2	Q5NT95 PAROL	Q5nt95 paralichthy
86	85.7	1439	2	Q97406 HALDI	Q97406 halloctis di
87	85.7	1458	2	Q910B9 ONCMY	Q910b9 oncorhynch
88	85.7	1463	2	Q59F89 HUMAN	Q59f89 homo sapien
89	85.7	1478	2	Q5AQJ3 EMENT	Q5aqj3 aspergillus
90	85.7	1546	2	Q8TER5 HUMAN	Q8ter5 homo sapien
91	85.7	1580	2	Q8BLX7 MOUSE	Q8blx7 mus musculus
92	85.7	1603	1	COGAL HUMAN	Q07032 homo sapien
93	85.7	1631	2	Q597Q0 CANFA	Q597q0 canis famil
94	85.7	1658	2	Q59GD4 HUMAN	Q59gd4 homo sapien
95	85.7	1745	1	C05A3 MOUSE	P25940 homo sapien
96	85.7	1804	2	Q80WR4 MOUSE	Q80wr4 mus musculus
97	85.7	1804	2	Q80WR4 MOUSE	Q80wr4 mus musculus
98	85.7	1806	1	COBA1 HUMAN	P12107 homo sapien
99	85.7	1806	2	Q5VT31 HUMAN	Q5vt31 homo sapien
100	85.7	2936	2	Q7YK88 CANFA	Q7yrk8 canis famil
101	83.3	132	2	P78429 HUMAN	P78429 homo sapien
102	83.3	165	2	Q73PN7 TREDE	Q73pn7 treponema d
103	83.3	242	1	BUTA_STAAM	Q5ga36 grouper iri
104	83.3	258	1	BUTA_STAAM	P66775 staphylococ

105	35	83.3	258	1	BUTA STAAN	P99120	staphylococ	178	34	81.0	105	2	Q984C8	RHILO	Q984c8	rhizobium 1
106	35	83.3	258	1	BUTA STAAR	Q6gkh9	staphylococ	179	34	81.0	123	2	Q67WN4	ORISA	Q67wn4	oryza sativ
107	35	83.3	258	1	BUTA STAAS	Q6czz8	staphylococ	180	34	81.0	145	2	Q61248	MOUSE	Q61248	mus musculus
108	35	83.3	258	1	BUTA STAAS	P66776	staphylococ	181	34	81.0	155	2	Q8CTK5	MOUSE	Q8ctk5	mus musculus
109	35	83.3	258	1	QSHJ22 STAAC	Q5hjp2	staphylococ	182	34	81.0	167	2	Q9TT85	PIG	Q9tt85	sus scrofa
110	35	83.3	299	1	COL34 CAEEL	P34687	caenorhabdi	183	34	81.0	184	2	Q92416	CAVPO	Q92416	cavia porce
111	35	83.3	299	2	Q61UE9 CAEBR	O61ue9	caenorhabdi	184	34	81.0	187	2	Q9BT15	HUMAN	Q9bt15	homo sapien
112	35	83.3	314	2	O17035 CAEEL	O17035	caenorhabdi	185	34	81.0	187	2	Q4VXW0	HUMAN	Q4vxw0	homo sapien
113	35	83.3	325	1	C09A1 RAT	P20850	rattus norv	186	34	81.0	193	2	Q4SAF1	TETNG	Q4saf1	tetraodon n
114	35	83.3	338	2	Q8BJU6 MOUSE	Q8bjue6	mus musculus	187	34	81.0	198	2	Q5JPA6	HUMAN	Q5jpa6	homo sapien
115	35	83.3	358	2	Q23222 CAEEL	Q23222	caenorhabdi	188	34	81.0	206	2	Q9RWY4	DEIRA	Q9rwy4	deinococcus
116	35	83.3	373	2	Q581V9 9TRYP	Q581v9	trypanosoma	189	34	81.0	209	2	Q8N811	HUMAN	Q8n811	homo sapien
117	35	83.3	398	2	Q60ZN4 CAEBR	O60zn4	caenorhabdi	190	34	81.0	222	2	Q5JMG2	ORISA	Q5jmg2	oryza sativ
118	35	83.3	393	2	Q9N3X9 CAEEL	O9n3x9	caenorhabdi	191	34	81.0	242	2	Q5YFM0	9VIRU	Q5yfm0	singapore g
119	35	83.3	494	2	Q7ZTK1 XENLA	Q7ztk1	xenopus lae	192	34	81.0	242	2	P89775	9HIVU	P89775	human immun
120	35	83.3	519	2	Q5B7L7 EMENI	Q5b7l7	aspergillus	193	34	81.0	244	2	Q9UKK4	MOUSE	Q9ukk4	mus musculus
121	35	83.3	538	2	Q53QO1 HUMAN	Q53qq1	homo sapien	194	34	81.0	253	2	Q5TPD7	ANOGA	Q5tpd7	anopheles g
122	35	83.3	568	2	Q6NUV1 BRARE	Q6nuv1	brachydanio	195	34	81.0	271	2	Q6P5Q3	HUMAN	Q6p5q3	homo sapien
123	35	83.3	636	1	C03A1 RAT	P13941	rattus norv	196	34	81.0	289	2	Q5LV74	SILPO	Q5lv74	silicibacte
124	35	83.3	643	2	Q6DFL1 XENLA	Q6dfll1	xenopus lae	197	34	81.0	294	1	GNMT	HUMAN	Q14749	homo sapien
125	35	83.3	675	1	C09A3 CHICK	P32017	gallus gall	198	34	81.0	295	2	Q5T8W2	HUMAN	Q5t8w2	homo sapien
126	35	83.3	675	2	Q90800 CHICK	Q90800	gallus gall	199	34	81.0	295	2	Q61K13	CAEBR	Q61k13	caenorhabdi
127	35	83.3	680	2	Q9D0D2 MOUSE	Q9d0d2	m mus muscu	200	34	81.0	297	2	Q6A6U9	PROAC	Q6a6u9	propionibac
128	35	83.3	697	2	Q5YZY6 NOCFA	O5zyz6	nocardia fa	201	34	81.0	314	2	Q4S325	TETNG	Q4s325	tetraodon n
129	35	83.3	733	2	Q9XZR2 CIOIN	O9xzz2	ciona intes	202	34	81.0	352	2	Q4RVB4	TETNG	Q4rvb4	tetraodon n
130	35	83.3	733	2	Q4H3C7 CIOIN	Q4h3c7	ciona intes	203	34	81.0	362	2	Q4SPT1	TETNG	Q4spt1	tetraodon n
131	35	83.3	786	2	Q8S6S2 9CAUD	Q8s6s2	mycobacteri	204	34	81.0	375	2	Q5TMG9	ANOGA	Q5tmg9	anopheles g
132	35	83.3	825	1	ICP0 HHV2H	P28284	human herpe	205	34	81.0	379	1	PSII	SCHPO	Q09912	schizosacch
133	35	83.3	826	2	P89473 HHV2	P89473	human herpe	206	34	81.0	380	2	Q8H8F3	ORISA	Q8h8f3	oryza sativ
134	35	83.3	832	2	Q96JF7 HUMAN	Q96jf7	homo sapien	207	34	81.0	380	2	Q98B41	RHILO	Q98b41	rhizobium 1
135	35	83.3	837	2	Q5H0Q6 XANOR	O5h0q6	xanthomonas	208	34	81.0	383	2	Q8K2L1	MOUSE	Q8k2l1	mus musculus
136	35	83.3	857	2	Q9Z883 MOUSE	Q9z883	mus musculus	209	34	81.0	390	2	Q4SQ42	TETNG	Q4sq42	tetraodon n
137	35	83.3	888	2	Q90796 CHICK	Q90796	gallus gall	210	34	81.0	392	2	Q8SZP7	DROME	Q8szp7	drosophila
138	35	83.3	894	2	Q3RC94 PONPY	Q3rc94	pongo pygma	211	34	81.0	401	2	Q4SB11	TETNG	Q4sb11	tetraodon n
139	35	83.3	905	2	Q5B0P8 EMENI	Q5b0p8	aspergillus	212	34	81.0	414	2	Q61SL1	CAEBR	Q61sl1	caenorhabdi
140	35	83.3	907	2	Q26312 STRPU	Q26312	strongyloce	213	34	81.0	467	2	Q727B9	DESVH	Q727b9	desulfovibr
141	35	83.3	921	1	C09A1 MOUSE	O05722	mus musculus	214	34	81.0	469	2	Q70598	RAT	Q70598	rattus norv
142	35	83.3	921	2	Q8BSQ1 MOUSE	Q8bsq4	mus musculus	215	34	81.0	494	2	Q58DC2	BOVIN	Q58dc2	bos taurus
143	35	83.3	966	1	N8R1 HUMAN	Q14596	homo sapien	216	34	81.0	500	2	Q627E1	CAEBR	Q627e1	caenorhabdi
144	35	83.3	966	2	Q5J7Q8 HUMAN	Q5j7q8	homo sapien	217	34	81.0	501	2	Q8CBW4	MOUSE	Q8cbw4	mus musculus
145	35	83.3	1045	2	Q801S8 XENLA	Q801s8	xenopus lae	218	34	81.0	501	2	Q8CAL3	MOUSE	Q8cal3	mus musculus
146	35	83.3	1046	2	Q8K388 MOUSE	Q8k388	mus musculus	219	34	81.0	501	2	Q8CA30	MOUSE	Q8ca30	mus musculus
147	35	83.3	1049	1	C03A1 BOVIN	P04258	bos taurus	220	34	81.0	501	2	Q8C8E2	MOUSE	Q8c8e2	mus musculus
148	35	83.3	1128	2	Q5F488 CHICK	P04258	bos taurus	221	34	81.0	501	2	Q6NV47	MOUSE	Q6nv47	mus musculus
149	35	83.3	1163	2	Q8N6U4 HUMAN	Q8n6u4	homo sapien	222	34	81.0	513	2	Q52237	MYXXA	Q52237	myxococcus
150	35	83.3	1191	2	Q9RHV0 STRRO	Q9rhv0	streptomyce	223	34	81.0	520	1	MARCO	HUMAN	Q9uew3	homo sapien
151	35	83.3	1225	2	Q6PCL3 MOUSE	Q6pcl3	mus musculus	224	34	81.0	520	2	Q4ZG40	HUMAN	Q4zg40	homo sapien
152	35	83.3	1306	2	Q7D9L6 MYCTU	Q7d9l6	mycobacteri	225	34	81.0	535	2	Q5RCJ6	PONPY	Q5rcj6	pongo pygma
153	35	83.3	1306	2	Q6MX28 MYCTU	O6mx28	mycobacteri	226	34	81.0	546	2	Q6P7U1	MOUSE	Q6p7u1	mus musculus
154	35	83.3	1306	2	Q7UIQ7 MYCBO	Q7ulq7	mycobacteri	227	34	81.0	551	2	Q5P1M3	AZOSE	Q5p1m3	azocarcus sp
155	35	83.3	1445	2	Q93251 RANCA	O93251	rana catesb	228	34	81.0	561	2	Q96IF4	HUMAN	Q96if4	homo sapien
156	35	83.3	1447	2	Q5NT96 PAROL	Q5nt96	paralichthy	229	34	81.0	565	2	Q4P1L4	USTMA	Q4p1l4	utilago ma
157	35	83.3	1447	2	Q91B91 XENLA	P02457	gallus gall	230	34	81.0	571	2	Q8N200	HUMAN	Q8n200	homo sapien
158	35	83.3	1453	1	C01A1 CHICK	P02457	gallus gall	231	34	81.0	583	2	Q9RKB1	STRCO	Q9rkb1	streptomyce
159	35	83.3	1463	2	Q5PQTE RAT	Q5pqt6	rattus norv	232	34	81.0	629	2	Q82RP5	STRAW	Q82rp5	streptomyce
160	35	83.3	1463	2	Q4W6W6 9CHON	Q4w6w6	raja kenoje	233	34	81.0	638	2	Q7PM40	ANOGA	Q7pm40	anopheles g
161	35	83.3	1466	1	C03A1 HUMAN	P02461	homo sapien	234	34	81.0	662	2	Q8VHY3	MOUSE	Q8vhy3	mus musculus
162	35	83.3	1466	2	Q53S91 HUMAN	Q53se91	homo sapien	235	34	81.0	674	2	Q7LGG2	XENLA	Q7lgg2	homo sapien
163	35	83.3	1466	2	Q541P8 HUMAN	Q541p8	homo sapien	236	34	81.0	676	2	Q6DE50	XENLA	Q6de50	xenopus lae
164	35	83.3	1670	1	C04A3 HUMAN	O01955	homo sapien	237	34	81.0	679	2	Q96RU3	HUMAN	Q96ru3	homo sapien
165	35	83.3	1682	2	Q9QZR5 MOUSE	Q9qzzr9	mus musculus	238	34	81.0	680	2	Q8BSX1	MOUSE	Q8bsx1	mus musculus
166	35	83.3	1752	2	Q07265 STRPU	Q07265	strongyloce	239	34	81.0	684	1	C09A3 HUMAN		Q14050	homo sapien
167	35	83.3	1787	2	Q9M4X9 CHLRE	Q9m4x9	chlamydomon	240	34	81.0	686	2	Q58EW9	XENLA	Q58ew9	xenopus lae
168	35	83.3	1845	2	Q80UA8 MOUSE	Q80ua8	mus musculus	241	34	81.0	689	1	C09A2 HUMAN		Q14055	homo sapien
169	35	83.3	1845	2	Q5ONQ9 MOUSE	Q5ong9	mus musculus	242	34	81.0	699	1	C08A2	MOUSE	P23318	mus musculus
170	35	83.3	1855	2	Q80ZF0 RAT	Q80zf0	rattus norv	243	34	81.0	715	2	Q4T9V1	TETNG	Q4t9v1	tetraodon n
171	35	83.3	1860	2	Q81ZC5 HUMAN	O81zcc6	homo sapien	244	34	81.0	718	2	Q4TA13	TETNG	Q4ta13	tetraodon n
172	35	83.3	2274	2	Q9Z1K7 MOUSE	Q9z1k7	mus musculus	245	34	81.0	747	1	C02A1 BOVIN		P02459	bos taurus
173	35	83.3	3295	2	Q83X71 STRRO	Q83x71	streptomyce	246	34	81.0	771	2	C080384	ORISA	O80384	oryza sativ
174	35	83.3	3651	2	Q83X69 STRRO	Q83x69	streptomyce	247	34	81.0	771	2	P93409	ORISA	P93409	oryza sativ
175	35	83.3	4226	2	Q8TKD6 METAC	Q8tkd6	methanosarc	248	34	81.0	782	2	Q67WN5	ORISA	Q67wn5	oryza sativ
176	34	81.0	60	2	Q4RBN4 TETNG	Q4rbn4	tetraodon n	249	34	81.0	799	2	Q8BNS7	MOUSE	Q8bns7	mus musculus
177	34	81.0	99	2	Q4RDQ7 TETNG	Q4rdq7	tetraodon n	250	34	81.0	802	2	Q7PYX1	ANOGA	Q7pyx1	anopheles g

397	33	78.6	490	2	Q8PKA2_XANAC	Q8PKA2_xanthomonas
398	33	78.6	511	2	Q4RH5_TETNG	Q4rhx5 tetraodon n
399	33	78.6	515	2	Q8LR3_ORYSA	Q8lra3 oryza sativ
400	33	78.6	534	2	Q6ZUT6_HUMAN	Q6zut6 homo sapien
401	33	78.6	546	2	Q6OV23_CAEBR	Q6ov23 caenorhabdi
402	33	78.6	546	2	Q99K97_MOUSE	Q99k97 mus musculu
403	33	78.6	550	2	Q88R3_PSEPK	Q88r3 pseudomonas
404	33	78.6	553	2	Q4SC69_TETNG	Q4sc69 tetraodon n
405	33	78.6	580	2	Q6CG49_YARLI	Q6cg49 yarrowia li
406	33	78.6	609	2	Q8PFN7_XANAC	Q8pfm7 xanthomonas
407	33	78.6	620	1	MBRL_HUMAN	Q4zin3 homo sapien
408	33	78.6	630	2	Q9LZE2_ARATH	Q9lze2 arabidopsis
409	33	78.6	632	2	Q9N2N7_HMPU	Q9n2n7 hemientrot
410	33	78.6	647	2	Q4RT24_TETNG	Q4rt24 tetraodon n
411	33	78.6	671	1	COLA1_RAT	P02454 rattus norv
412	33	78.6	680	1	COA1_MOUSE	Q05306 mus musculu
413	33	78.6	681	2	Q6AZQ8_XENLA	Q6azq8 xenopus lae
414	33	78.6	702	2	Q92541_HUMAN	Q92541 homo sapien
415	33	78.6	708	2	Q7ZWN8_XENLA	Q7zwn8 xenopus lae
416	33	78.6	715	1	ATG13_YARLI	Q6c315 yarrowia li
417	33	78.6	717	2	Q6C9B8_YARLI	Q6c9b8 yarrowia li
418	33	78.6	730	2	Q26052_PARLI	Q26052 paracentrot
419	33	78.6	734	2	Q6IVJ4_CIOIN	Q6ivj4 ciona intes
420	33	78.6	757	2	Q8NB69_HUMAN	Q8nb69 homo sapien
421	33	78.6	757	2	Q5JV63_HUMAN	Q5jv63 homo sapien
422	33	78.6	771	2	Q6LDU2_CAEBR	Q6ldu2 caenorhabdi
423	33	78.6	789	2	Q4I8V2_GIBZE	Q4i8v2 gibberella
424	33	78.6	799	2	Q96N31_HUMAN	Q96n31 homo sapien
425	33	78.6	799	2	Q8A4J6_BACTN	Q8a4j6 bacteroides
426	33	78.6	802	1	Y0250_HUMAN	Q92540 homo sapien
427	33	78.6	802	2	Q6TV06_HUMAN	Q6tv06 homo sapien
428	33	78.6	809	2	Q93485_ONCMY	Q93485 oncorhynch
429	33	78.6	812	2	Q82N72_STRAW	Q82n72 streptomyce
430	33	78.6	830	2	Q7TT25_MOUSE	Q7tt25 mus musculu
431	33	78.6	839	2	Q7Z7H9_HUMAN	Q7z7h9 homo sapien
432	33	78.6	839	2	Q4UPE0_XANCP	Q4upe0 xanthomonas
433	33	78.6	839	2	Q8P3W4_XANCP	Q8p3w4 xanthomonas
434	33	78.6	854	2	Q8IVT9_HUMAN	Q8ivt9 homo sapien
435	33	78.6	866	2	Q8PFE7_XANAC	Q8pfe7 xanthomonas
436	33	78.6	879	2	Q99MS1_MOUSE	Q99ms1 mus musculu
437	33	78.6	879	2	Q8K2I4_MOUSE	Q8k2i4 mus musculu
438	33	78.6	888	2	Q5H5V9_XANOR	Q5h5v9 xanthomonas
439	33	78.6	894	2	Q8MW54_MVTGA	Q8mw54 mytilus gal
440	33	78.6	920	2	Q78EC6_9MURI	Q78ec6 mus sp. typ
441	33	78.6	932	2	Q4SV45_TETNG	Q4sv45 tetraodon n
442	33	78.6	971	2	Q4RJ71_TETNG	Q4rj71 tetraodon n
443	33	78.6	977	2	Q4SEP8_TETNG	Q4sep8 tetraodon n
444	33	78.6	994	2	Q4RYP8_TETNG	Q4ryp8 tetraodon n
445	33	78.6	1009	2	Q4SNW2_TETNG	Q4snw2 tetraodon n
446	33	78.6	1019	1	CO6A2_HUMAN	P12110 homo sapien
447	33	78.6	1019	2	Q8P0Q1_HUMAN	Q8p0q1 homo sapien
448	33	78.6	1029	1	CO6A2_MOUSE	Q02788 mus musculu
449	33	78.6	1034	2	Q8K229_MOUSE	Q8k229 mus musculu
450	33	78.6	1037	2	Q9QY01_MOUSE	Q9qy01 mus musculu
451	33	78.6	1037	2	Q8WTP4_MOUSE	Q8wtp4 mus musculu
452	33	78.6	1053	2	Q820S4_MOUSE	Q820s4 mus musculu
453	33	78.6	1056	2	Q4UWV2_XANCP	Q4uwv2 xanthomonas
454	33	78.6	1056	2	Q8P875_XANCP	Q8p875 xanthomonas
455	33	78.6	1056	2	Q80TV7_MOUSE	Q80tv7 mus musculu
456	33	78.6	1064	2	Q8DIP1_SYNEL	Q8dip1 synechococc
457	33	78.6	1068	1	H1P1_HUMAN	Q75146 homo sapien
458	33	78.6	1068	2	Q5RD46_PONPY	Q5rd46 pongo pygma
459	33	78.6	1071	2	Q8IXC1_HUMAN	Q8ixc1 homo sapien
460	33	78.6	1084	2	Q53R14_HUMAN	Q53r14 homo sapien
461	33	78.6	1097	2	Q4RHF1_TETNG	Q4rhf1 tetraodon n
462	33	78.6	1099	2	Q5JV62_HUMAN	Q5jv62 homo sapien
463	33	78.6	1118	2	Q5SUB4_MOUSE	Q5sub4 mus musculu
464	33	78.6	1137	2	Q8IXC2_HUMAN	Q8ixc2 homo sapien
465	33	78.6	1137	2	Q5TIQ0_HUMAN	Q5tiq0 homo sapien
466	33	78.6	1178	2	Q6PIE0_HUMAN	Q6pie0 homo sapien
467	33	78.6	1262	1	CO3A1_CHICK	P12105 gallus gall
468	33	78.6	1307	2	Q8JFF7_XENLA	Q8jff7 xenopus lae
469	33	78.6	1346	2	Q8UUJ3_ONCKE	Q8uuj3 oncorhynch

470	33	78.6	1352	2	Q90YJ0_BRARE	Q90yj0 brachydanio
471	33	78.6	1352	2	Q6IOX2_BRARE	Q6iox2 brachydanio
472	33	78.6	1356	1	COLA2_ONCMY	Q91484 oncorhynch
473	33	78.6	1367	1	COPAL_MOUSE	Q35206 mus musculu
474	33	78.6	1380	2	Q4S053_TETNG	Q4s053 tetraodon n
475	33	78.6	1403	2	Q4RKG5_TETNG	Q4rkg5 tetraodon n
476	33	78.6	1449	2	Q910C0_ONCMY	Q910c0 oncorhynch
477	33	78.6	1468	2	Q4S014_TETNG	Q4s014 tetraodon n
478	33	78.6	1469	2	Q6PD04_MOUSE	Q6pd04 mus musculu
479	33	78.6	1517	2	Q6KAS1_MOUSE	Q6kas1 mus musculu
480	33	78.6	1599	2	Q6V1P0_9ACTO	Q6v1p0 streptomyce
481	33	78.6	1626	2	Q8NFW1_HUMAN	Q8nfw1 homo sapien
482	33	78.6	1655	2	Q4SDF0_TETNG	Q4sdf0 tetraodon n
483	33	78.6	1691	2	Q9ES01_MOUSE	Q9esq1 mus musculu
484	33	78.6	1705	2	Q5VZA9_HUMAN	Q5vza9 homo sapien
485	33	78.6	1708	2	Q7PVR6_9NOGA	Q7pvr6 anopheles g
486	33	78.6	1712	1	CO4A2_HUMAN	P08572 homo sapien
487	33	78.6	1736	1	COBA2_MOUSE	Q64739 mus musculu
488	33	78.6	1737	2	Q9J104_RAT	Q9j104 rattus norv
489	33	78.6	1758	1	CO4A2_CAEL	P17140 caenorhabdi
490	33	78.6	1758	2	Q61315_CAEBR	Q61315 caenorhabdi
491	33	78.6	1759	1	CO4A1_CAEBR	P17139 caenorhabdi
492	33	78.6	1759	2	Q611Q5_CAEBR	Q611q5 caenorhabdi
493	33	78.6	1827	2	Q8UUM5_ORYLA	Q8uum5 oryzias lat
494	33	78.6	1877	2	Q4RWT3_TETNG	Q4rwt3 tetraodon n
495	33	78.6	1888	1	COEAL_CHICK	P32018 gallus gall
496	33	78.6	2124	2	Q6GVP6_MYCAV	Q6gvp6 mycobacteri
497	33	78.6	2126	2	Q7VEV1_MYCBO	Q7rev1 mycobacteri
498	33	78.6	2126	2	P94996_MYCTU	P94996 mycobacteri
499	33	78.6	2689	2	Q4QF64_LEIMA	Q4qf64 leishmania
500	33	78.6	3528	2	Q5Y9G4_9ACTO	Q5y9g4 aeromicrobi

ALIGNMENTS

RESULT 1					
Q99228_HUMAN PRELIMINARY;					PRT; 584 AA.
ID	Q99228_HUMAN	PRELIMINARY;	PRT;	584 AA.	
AC	Q99228;				
DT	01-NOV-1996	(TrEMBLrel. 01, Created)			
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)			
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)			
DE	Alpha-1 type XIII collagen.				
GN	Name=COL13A1;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;				
OC	Homo.				
OX	NCBI_TaxID=9606;				
RN	[1] NUCLEOTIDE SEQUENCE.				
RP	MEDLINE=91009112; PubMed=1698771;				
RX	Pihlajaniemi T., Tamminen M.;				
RA	"The alpha 1 chain of type XIII collagen consists of three collagenous				
RT	and four noncollagenous domains, and its primary transcript undergoes				
RT	complex alternative splicing.";				
RL	J. Biol. Chem. 265:16922-16928(1990).				
RL	EMBL; M59217; AAA51685.1; -; mRNA.				
DR	Ensembl; ENSG00000197467; Homo sapiens.				
DR	GO; GO:0005737; C:cytoplasm; IEA.				
DR	GO; GO:0005515; F:protein binding; IEA.				
DR	GO; GO:0005198; F:structural molecule activity; IEA.				
DR	GO; GO:0007155; F:cell adhesion; IEA.				
DR	GO; GO:0006817; P:phosphate transport; IEA.				
DR	InterPro; IPR008161; C1g helix.				
DR	InterPro; IPR008160; Collagen.				
DR	Pfam; PF01391; Collagen; 8				
DR	ProDom; PD000007; C1g_helix; 3.				
KW	Collagen.				
SQ	SEQUENCE 584 AA; 56522 MW; 090A5BA4F007BF4B CRC64;				
Query Match					100.0%; Score 42; DB 2; Length 584;

Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
| | | | | | | |
Db 251 GAEGSPGL 258

RESULT 2
Q14035 HUMAN
ID Q14035 HUMAN PRELIMINARY; PRT; 623 AA.
AC Q14035
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha-1 type XIII collagen.
GN Name=COL13A1;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=91373404; PubMed=1894651;
RA Tikka L., Elomaa O., Pihlajaniemi T., Tryggvason K.;
RT "Human alpha 1 (XIII) collagen gene. Multiple forms of the gene
transcripts are generated through complex alternative splicing of
several short exons";
RT J. Biol. Chem. 266:17713-17719(1991).
RL EMBL; M69008; AA52754.1; -, Genomic DNA.
DR EMBL; M69884; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68985; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68986; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68987; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68988; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68989; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68990; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68991; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68992; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68993; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68994; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68995; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68996; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68997; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68998; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68999; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69000; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69001; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69002; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69003; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69004; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69005; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69006; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69007; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69009; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69010; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68978; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68977; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68976; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68975; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M68974; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M69010; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M81617; AA52754.1; JOINED; Genomic DNA.
DR EMBL; M81618; AA52754.1; JOINED; Genomic DNA.
DR Ensembl; ENSG0000197467; Homo sapiens.
DR GO; GO:0005737; Cytoplasm; IEA.
DR CO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g helix.

DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; C1g_helix; 3.
KW Collagen.
SQ SEQUENCE 623 AA; 60470 MW; 887350BCE6958692 CRC64;

Query Match 100.0%; Score 42; DB 2; Length 623;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
| | | | | | | |
Db 263 GAEGSPGL 270

RESULT 3
Q5TAT5 HUMAN
ID Q5TAT5 HUMAN PRELIMINARY; PRT; 652 AA.
AC Q5TAT5
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Collagen, type XIII, alpha 1 (Fragment).
GN Name=COL13A1; ORFNames=RP11-26212.1-002;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lawlor S.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL113892; CA115451.1; -, Genomic DNA.
DR EMBL; AC024601; CA115451.1; JOINED; Genomic DNA.
DR EMBL; AC025426; CA115451.1; JOINED; Genomic DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; C1g_helix; 3.
KW Collagen.
FT NON_TER
SQ SEQUENCE 652 AA; 63277 MW; BC950B456DBF34A6 CRC64;

Query Match 100.0%; Score 42; DB 2; Length 652;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
| | | | | | | |
Db 291 GAEGSPGL 298

RESULT 4
Q5TAT4 HUMAN
ID Q5TAT4 HUMAN PRELIMINARY; PRT; 683 AA.
AC Q5TAT4;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Collagen, type XIII, alpha 1.
GN Name=COL13A1; ORFNames=RP11-26212.1-004;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lawlor S.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL138925; CAIL15452.1; -; Genomic DNA.
 DR EMBL; AC024601; CAIL15452.1; JOINED; Genomic DNA.
 DR EMBL; AC025426; CAIL15452.1; JOINED; Genomic DNA.
 DR Ensembl; ENSG00000197467; Homo sapiens.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 2.
 KW Collagen.
 SQ SEQUENCE 683 AA; 66377 MW; 3BB550FD03062C30 CRC64;

Query Match 100.0%; Score 42; DB 2; Length 683;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 |||||
 Db 334 GAEGSPGL 341

RESULT 5

Q9NQ52 HUMAN
 ID Q9NQ52_HUMAN PRELIMINARY; PRT; 717 AA.
 AC Q9NQ52;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Type XIII collagen.
 GN Name=COL13A1; Synonyms=COLXIIIA1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Snellman A., Tu H., Vaeisaenen T., Kvist A.P., Huhtala P.,
 RA Pihlajaniemi T.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ293624; CAC00688.1; -; mRNA.
 DR Ensembl; ENSG00000197467; Homo sapiens.
 DR HGNC; HGNC:21190; COL13A1.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 717 AA; 69964 MW; A311E9C7D3E87577 CRC64;

Query Match 100.0%; Score 42; DB 2; Length 717;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 |||||
 Db 356 GAEGSPGL 363

RESULT 6

Q5TAT6 HUMAN
 ID Q5TAT6_HUMAN PRELIMINARY; PRT; 717 AA.
 AC Q5TAT6;
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Collagen, type XIII, alpha 1.

GN Name=COL13A1; ORFNames=RP11-26212.1-001;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Lawlor S.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL138925; CAIL15450.1; -; Genomic DNA.
 DR EMBL; AC024601; CAIL15450.1; JOINED; Genomic DNA.
 DR EMBL; AC025426; CAIL15450.1; JOINED; Genomic DNA.
 DR HGNC; HGNC:21190; COL13A1.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 717 AA; 69950 MW; FD12CA80CC93540F CRC64;

Query Match 100.0%; Score 42; DB 2; Length 717;
 Best Local Similarity 100.0%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 |||||
 Db 356 GAEGSPGL 363

RESULT 7

Q4RZT4 TETNG
 ID Q4RZT4_TETNG PRELIMINARY; PRT; 405 AA.
 AC Q4RZT4;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome 18 SCAP14786, whole genome shotgun sequence.
 GN ORFNames=GSTENG00026335001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Farra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RA Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAAE01014786; CAG06098.1; -; Genomic DNA.
 SQ SEQUENCE 405 AA; 44006 MW; 852BD5D6D09297F0 CRC64;

Query Match 92.9%; Score 39; DB 2; Length 405;
 Best Local Similarity 87.5%; Pred. No. 90;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAEGSPGL 8
 ||:|||||
 Db 54 GADGSPGL 61

RESULT 8

Q4SB07 TETNG
 ID Q4SB07 TETNG PRELIMINARY; PRT; 856 AA.
 AC Q4SB07;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DE Chromosome undetermined SCAFI4677, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG0021178001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAAG01014677; CAG02175.1; -; Genomic_DNA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 1.
 DR Pfam; PF01391; Collagen; 8.
 DR ProDom; PD000007; Clg_helix; 3.
 DR ProDom; PD003923; Procollagn4_C; 1.
 DR SMART; SM00111; C4; 1.
 KW Collagen.
 FT NON_TER
 SQ SEQUENCE 856 AA; 84113 MW; A5FAC617906E701A CRC64;

Query Match 92.9%; Score 39; DB 2; Length 856;
 Best Local Similarity 87.5%; Pred. No. 1.9e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAEGSPGL 8
 ||:|||||
 Db 285 GAEGSPGV 292

RESULT 9

Q4SZ73 TETNG
 ID Q4SZ73 TETNG PRELIMINARY; PRT; 1026 AA.
 AC Q4SZ73;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Chromosome undetermined SCAFI1805, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG0009982001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
 RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Winkler P., Lander E.S., Weissbach J., Roest Crolius H.,
 RA "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 RT the early vertebrate proto-karyotype."
 RL Nature 431:946-957(2004).
 RN [2]

RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; CAAG01011805; CAF94059.1; -; Genomic_DNA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR001442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 10.
 DR ProDom; PD000007; Clg_helix; 4.
 DR ProDom; PD003923; Procollagn4_C; 2.
 DR SMART; SM00111; C4; 2.
 KW Collagen.
 FT NON_TER
 SQ SEQUENCE 1026 AA; 103855 MW; 1D61B3FF1779C267 CRC64;

Query Match 92.9%; Score 39; DB 2; Length 1026;
 Best Local Similarity 87.5%; Pred. No. 2.3e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAEGSPGL 8
 ||:|||||
 Db 454 GADGSPGL 461

RESULT 10
 CO4A2 MOUSE
 ID CO4A2 MOUSE STANDARD; PRT; 1707 AA.
 AC P08122; Q61375;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Collagen alpha 2(IV) chain precursor.
 GN Name=Col4a2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;


```
Db      1135 GAQSPGL 1142

RESULT 11
Q5BW21 SCHJA PRELIMINARY; PRT; 164 AA.
AC Q5BW21;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
OS Schistosoma japonicum (blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Han Z.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY812515; AAX28404.1; -; mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 2.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 164 AA; 15488 MW; CB4E5C07BBFE588D CRC64;

Query Match 90.5%; Score 38; DB 2; Length 164;
Best Local Similarity 87.5%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAQSPGL 8
Db 134 GAQSPGL 141

RESULT 12
Q810M6 MOUSE
ID Q810M6 MOUSE PRELIMINARY; PRT; 228 AA.
AC Q810M6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein MGC58818 (Novel protein).
GN Name=MGC58818; Synonyms=RP23-9619.2; ORFNames=RP23-9619.2-001;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Kjaerberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uddin T.B., Toohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Grafham D.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC049758; AHA49758.1; -; mRNA.
DR EMBL; AL645589; CA126108.1; -; Genomic DNA.
DR Ensembl; ENSMUSG00000053783; Mus musculus.
KW Hypothetical protein.
SQ SEQUENCE 228 AA; 25512 MW; 0DDABD34E39DB504 CRC64;

Query Match 90.5%; Score 38; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAQSPG 7
Db 28 GAQSPG 34

RESULT 13
Q9R149 CAVPO
ID Q9R149_CAVPO PRELIMINARY; PRT; 230 AA.
AC Q9R149;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Pro-alpha-1 type I collagen (Fragment).
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
OC Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=Hartley; TISSUE=Trachea;
RA Morishima Y., Uchida Y., Nomura A., Ishii Y., Sakamoto T.,
RA Sekizawa K.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF169346; AAD49346.1; -; mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005515; P:protein binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD0000007; Clg_helix; 1.
KW Collagen.
FT NON_TER 1
FT NON_TER 230
SQ SEQUENCE 230 AA; 20425 MW; 1A465F92779D9A71 CRC64;

Query Match 90.5%; Score 38; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAQSPG 7
Db 150 GAQSPG 156

RESULT 14
QSR949 PONPY
ID QSR949 PONPY PRELIMINARY; PRT; 460 AA.
AC QSR949;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
```

DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Hypothetical protein DKFZp459P0517.
GN Name=DKFZp459P0517;
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominoidea;
OC Pongo.
OX NCBI_TaxID=9600;
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Cortex;
RG The German cDNA Consortium;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: The muscarinic acetylcholine receptor mediates various
CC cellular responses, including inhibition of adenylate cyclase,
CC breakdown of phosphoinositides and modulation of potassium
CC channels through the action of G proteins. Primary transducing
CC effect is PI turnover (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the G-protein coupled receptor 1 family.
CC EMBL; CR859546; CA91711.1; -; mRNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0045211; C:postsynaptic membrane; IEA.
DR GO; GO:0004981; F:muscarinic acetylcholine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR InterPro; IPR000995; MusAcC_receptor.
DR InterPro; IPR002228; MusacM1_receptor.
DR PANTHER; PTHR19266:SF86; MusacM1_receptor; 1.
DR Pfam; PF00001; 7cm.1; 1.
DR PRINTS; PR00237; GPCR_Rhodopsn.
DR PRINTS; PR00243; MUSCARINICR.
DR PRINTS; PR00538; MUSCARINICMR.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL2_1; 1.
KW G-protein coupled receptor; Hypothetical protein;
KW Postsynaptic membrane; Receptor; Transducer; Transmembrane.
SQ SEQUENCE 460 AA; 51322 MW; 51095410807C0FFC CRC64;

Query Match 90.5%; Score 38; DB 2; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 247 GAEGSPG 253
|||||
RESULT 15
Q99LL6 MOUSE
ID Q99LL6 MOUSE PRELIMINARY; PRT; 589 AA.
AC Q99LL6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collal protein (Fragment).
GN Name=Collal;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N;
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
RC virgin mouse. Taken by biopsy;
RX MEDLINE=22388257; PubMed=1477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Colling F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=FVB/N;
RC TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
RC virgin mouse. Taken by biopsy;
RA Strausberg R.; to the EMBL/GenBank/DBJ databases.
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003198; AA030198.1; -; mRNA.
DR MGI; MGI:188467; Collal.
GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 589 AA; 58805 MW; 81847495E5E05CEP CRC64;

Query Match 90.5%; Score 38; DB 2; Length 589;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 144 GAEGSPG 150
|||||
RESULT 16
Q867Z8 BACRY
ID Q867Z8 BACRY PRELIMINARY; PRT; 698 AA.
AC Q867Z8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Scarlet.
GN Name=scarlet;
OS Bactrocera tryoni (Queensland fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Tephritidae; Tephritidae; Bactrocera; Bactrocera.
OX NCBI_TaxID=59916;
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RA Zhao J., Bennet C.L., Stewart G.J., Frommer M., Raphael K.A.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -!- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AY172185; AAO65145.1; -; Genomic_DNA.
DR EMBL; AY172186; AAO65146.1; -; mRNA.

```

DR GO: 0016020; C:membrane; IEA.
DR GO: 0005524; F:ATP binding; IEA.
DR GO: 0016887; F:ATPase activity; IEA.
DR GO: 0004262; F:ATPase activity; coupled to transmembrane m. . .; IEA.
DR GO: 0000016; F:nucleotide binding; IEA.
DR GO: 0015546; F:permease activity; IEA.
DR GO: 0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR000412; ABC 2.
DR InterPro: IPR003439; ABC_transp_like.
DR InterPro: IPR005284; Pigment_permease.
DR Pfam: PF01061; ABC2_membrane; 1.
DR Pfam: PF00005; ABC_tran; 1.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART: SM00382; AAA; 1.1204; 1.
DR TIGRFAMs: TIGR00955; 3a01204; 1.
DR PROSITE: PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE: PS00893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Nucleotide-binding; Transmembrane; Transport.
KW SEQUENCE 698 AA; 77286 MW; 1495D98D799EFD8A CRC64;
SQ
Query Match 90.5%; Score 38; DB 2; Length 698;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
Db 40 GAEGSPG 46
RESULT 17
COLAL_BOVIN
ID COLAL_BOVIN STANDARD; PRT; 779 AA.
AC P02453;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(I) chain (Fragments).
GN Name=COL1A1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP PROTEIN SEQUENCE OF 1-19.
RX MEDLINE=72255334; PubMed=4115172;
RA Rautenberg J., Timpl R., Furthmayr H.;
RT "Structural characterization of N-terminal antigenic determinants in
RT calf and human collagen.";
RL Eur. J. Biochem. 27:231-237 (1972).
RN [2]
RP PROTEIN SEQUENCE OF 20-145.
RX MEDLINE=76022320; PubMed=1164916;
RA Fietzek P.P., Kuehn K.;
RT "The covalent structure of collagen: amino-acid sequence of the
RT cyanogen-bromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5
RT from calf-skin collagen.";
RL Eur. J. Biochem. 52:77-82 (1975).
RN [3]
RP PROTEIN SEQUENCE OF 146-294.
RX MEDLINE=73049499; PubMed=4673951; DOI=10.1016/0014-5793(72)80545-3;
RA Fietzek P.P., Wendt P., Kell I., Kuehn K.;
RT "The covalent structure of collagen: amino acid sequence of alpha-1-
RT CB3 from calf skin collagen.";
RL FEBS Lett. 26:74-76 (1972).
RN [4]
RP PROTEIN SEQUENCE OF 235-562.
RX MEDLINE=74086118; PubMed=4359390;
RA Fietzek P.P., Rextrodt F.W., Hopper K.E., Kuehn K.;
RT "The covalent structure of collagen. 2. The amino-acid sequence of
RT alpha-1-CB7 from calf-skin collagen.";
RL Eur. J. Biochem. 39:396-400 (1973).
RN [5]
RP PROTEIN SEQUENCE OF 563-675.
RX MEDLINE=73042276; PubMed=4343808;
RA Wendt P., Mark K.V.D., Rextrodt F., Kuehn K.;
RT "The covalent structure of collagen. The amino-acid sequence of the
RT 112-residues. Amino-terminal part of peptide alpha-1-CB6 from calf-
RT skin collagen.";
RL Eur. J. Biochem. 30:169-183 (1972).
RN [6]
RP PROTEIN SEQUENCE OF 676-751.
RX MEDLINE=73042275; PubMed=4343807;
RA Fietzek P.P., Rextrodt F.W., Wendt P., Stark M., Kuehn K.;
RT "The covalent structure of collagen. Amino-acid sequence of peptide
RT alpha-1-CB6-C2.";
RL Eur. J. Biochem. 30:163-168 (1972).
RN [7]
RP PROTEIN SEQUENCE OF 759-779.
RX PubMed=11946479;
RA Rautenberg J., Fietzek P., Rextrodt F., Becker U., Stark M., Kuehn K.;
RT "The amino acid sequence of the carboxyterminal nonhelical cross link
RT region of the alpha 1 chain of calf skin collagen.";
RL FEBS Lett. 21:75-79 (1972).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -!- PTM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains. Pro-726 is the only 3-hydroxypro and the only hydroxylated
CC proline in position X.
CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxyl group.
CC -!- MISCELLANEOUS: The complete chain contains 1052 residues.
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC PIR; A91193; CGB01S.
CC InterPro: IPR008161; Clg_helix.
CC InterPro: IPR008160; Collagen.
CC InterPro: IPR001007; WVF_C.
CC Pfam: PF01391; Collagen; 12.
CC ProDom: PD000007; Clg_helix; 2.
CC PROSITE: PS01208; WVF_1; PARTIAL.
CC Collagen; Direct protein sequencing; Extracellular matrix;
CC Glycoprotein; Hydroxylation; Pyrrolidone carboxylic acid; Repeat;
KW Structural protein.
KW MOD_RES 1 1 Pyrrolidone carboxylic acid.
FT MOD_RES 9 9 Alllysine.
FT MOD_RES 103 103 5-hydroxylysine. (Potential).
FT MOD_RES 115 115 5-hydroxylysine (Potential).
FT MOD_RES 124 124 5-hydroxylysine (Potential).
FT MOD_RES 274 274 5-hydroxylysine (Potential).
FT MOD_RES 346 346 5-hydroxylysine (Potential).
FT MOD_RES 424 424 5-hydroxylysine (Potential).
FT MOD_RES 496 496 5-hydroxylysine (Potential).
FT MOD_RES 658 658 5-hydroxylysine (Potential).
FT MOD_RES 670 670 5-hydroxylysine (Potential).
FT MOD_RES 726 726 3-hydroxyproline.
FT CARBOHYD 103 103 O-linked (Gal. . .).
FT NON_CONS 145 146
SQ SEQUENCE 779 AA; 70346 MW; E554A7FF084283D1 CRC64;
Query Match 90.5%; Score 38; DB 1; Length 779;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 GAEGSPG 7
Db      581 GAEGSPG 587

RESULT 18
ZFY28 HUMAN          STANDARD;          PRT;      887 AA.
AC      Q9HCC9;
DT      05-JUL-2004 (Rel. 44, Last sequence update)
DE      13-SEP-2005 (Rel. 48, Last annotation update)
DE      Zinc finger FYVE domain containing protein 28.
GN      Name=ZFY28; Synonyms=KIA1643;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC      Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC      TISSUE=Brain;
RX      MEDLINE=20450683; PubMed=10997877;
RA      Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes.
RT      XVIII. The complete sequences of 100 new cDNA clones from brain which
RT      code for large proteins in vitro.";
RL      DNA Res. 7:273-281(2000).
CC      -1- SIMILARITY: Contains 1 FYVE-type zinc finger.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
DR      EMBL; AB046863; BAB13469.1; ALT_INIT; mRNA.
DR      Ensembl; ENSG00000159733; Homo sapiens.
DR      HGNC; HGNC:29334; ZFYVE28.
DR      InterPro; IPR000306; Znf_FYVE.
DR      Pfam; PF01363; FYVE; 1.
DR      SMART; SM00064; FYVE; 1.
DR      PROSITE; PS0178; ZF_FYVE; 1.
KW      Metal-binding; Zinc; Zinc-finger.
FT      ZN FING      817      879      FYVE-type.
SQ      SEQUENCE      887 AA; 96500 MW; B80E8E8A6103C6C4 CRC64;

Query Match      90.5%; Score 38; DB 1; Length 887;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      1 GAEGSPG 7
Db      374 GAEGSPG 380

RESULT 19
Q6LAN8 HUMAN
ID      Q6LAN8 HUMAN PRELIMINARY;          PRT;      1069 AA.
AC      Q6LAN8;
DT      05-JUL-2004 (TrEMBLrel. 27, Created)
DT      05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE      Collagen type I alpha 1 (Fragment).
GN      Name=COL1A1;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC      Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      MEDLINE=97141927; PubMed=8988177;

```

```

RA      Simon M.-P., Pedoutour F., Sirvent N., Grosgeorge J., Minoletti F.,
RA      Coindre J.-M., Terrier-Lacombe M.-J., Mandahl N., Craver R.D.,
RA      Blin N., Sozzi G., Turc-Carel C., O'Brien K.P., Redra D., Fransson I.,
RA      Guilbaud C., Dumanaki J.P.;
RT      "Deregulation of the platelet-derived growth factor B-chain gene via
RT      fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans
RT      and giant-cell fibroblastoma.";
RL      Nat. Genet. 15:95-98(1997).
RN      [2]
RP      NUCLEOTIDE SEQUENCE.
RA      O'Brien K.P.;
RA      Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RL      EMBL; X98705; CAA67261.1; -; Genomic_DNA.
DR      GO; GO:0005737; Cytoplasm; IEA.
DR      GO; GO:0005515; F:protein binding; IEA.
DR      GO; GO:0007155; P:cell adhesion; IEA.
DR      GO; GO:0006817; P:phosphate transport; IEA.
DR      InterPro; IPR008161; Clg_helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR001007; VWF_C.
DR      Pfam; PF01391; Collagen; 16.
DR      Pfam; PF00093; VWC; 1.
DR      ProDom; PD000007; Clg_helix; 3.
DR      SMART; SM00214; VWC; 1.
DR      PROSITE; PS01208; VWF_C_1; 1.
DR      PROSITE; PS50184; VWF_C_2; 1.
KW      Collagen.
FT      NON TER      1069      1069
SQ      SEQUENCE      1069 AA; 97445 MW; EE279B10572FB980 CRC64;

Query Match      90.5%; Score 38; DB 2; Length 1069;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches      7; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

Qy      1 GAEGSPG 7
Db      1019 GAEGSPG 1025

RESULT 20
COL1A1 MOUSE
ID      COL1A1 MOUSE          STANDARD;          PRT;      1453 AA.
AC      P11087; O53WT0; Q60635; Q61367; Q61427; Q63919;
DT      01-JUL-1989 (Rel. 11, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      13-SEP-2005 (Rel. 48, Last annotation update)
DE      Collagen alpha 1(I) chain precursor.
DE      Name=Col1a1; Synonyms=Colal;
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC      Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RC      STRAIN=FVE/N.
RX      MEDLINE=96033240; PubMed=8535610;
RA      Li S.W., Khillan J., Prockop D.J.;
RT      "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
RT      of type I procollagen.";
RL      Matrix Biol. 14:593-595(1995).
RN      [2]
RP      NUCLEOTIDE SEQUENCE OF 1-185 AND 1030-1453.
RX      MEDLINE=94092741; PubMed=8268229; DOI=10.1016/0167-4781(93)90016-7;
RA      Fenton S.P., Lamande S.R., Hannagan M., Stacey A., Jaenisch R.,
RA      Bateman J.F.;
RT      "Genomic sequence of mouse COL1A1 encoding the collagen propeptides.";
RL      Biochim. Biophys. Acta 1216:469-474(1993).
RN      [3]
RP      NUCLEOTIDE SEQUENCE OF 1-942.
RC      STRAIN=C57BL/6; TISSUE=Liver;
RX      MEDLINE=94344105; PubMed=8065328;
RA      Rhodes K., Rippe R.A., Umezawa A., Nehls M., Brenner D.A., Breindl M.;

```

RT "DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect mechanism.";
 RL Mol. Cell. Biol. 14:5950-5960(1994).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 518-1128.
 RX MEDLINE=86137403; PubMed=3841523; DOI=10.1016/0378-1119(85)90329-4;
 RA French B.T., Lee W.-H., Maul G.G.;
 RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen protein.";
 RL Gene 39:311-312(1985).
 RN [5]
 RP NUCLEOTIDE SEQUENCE OF 735-1130.
 RX MEDLINE=83141374; PubMed=6298597;
 RA Monson J.M., Friedman J., McCarthy B.J.;
 RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene: evidence for a mouse B1 element within the gene.";
 RL Mol. Cell. Biol. 2:1362-1371(1982).
 RN [6]
 RP NUCLEOTIDE SEQUENCE OF 735-878 AND 1005-1058.
 RX MEDLINE=83157109; PubMed=6219867;
 RA Monson J.M., McCarthy B.J.;
 RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene: evidence for insertions or deletions in gene coding sequences.";
 RL DNA 1:59-69(1981).
 RN [7]
 RP NUCLEOTIDE SEQUENCE OF 1-25.
 RX MEDLINE=84170331; PubMed=6324198;
 RA Harbers K., Kuehn M., Delius H., Jaenisch R.;
 RT "Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads to embryonic lethal mutation in mice.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1504-1508(1984).
 RN [8]
 RP NUCLEOTIDE SEQUENCE OF 1442-1453.
 RX MEDLINE=88124276; PubMed=3340560;
 RA Mooslehner K., Harbers K.;
 RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of the 3'-untranslated region.";
 RL Nucleic Acids Res. 16:773-773(1988).
 RN [9]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1442-1453.
 RX PubMed=2054384;
 RA Metaeranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and bones. In bones the fibrils are mineralized with calcium hydroxyapatite.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -!- SIMILARITY: Contains 1 WFCC domain.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous gene model prediction.
 CC -----
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
 CC -----
 DR EMBL; U08020; AAB88912.1; -; mRNA.
 DR EMBL; S67530; AAB29424.1; -; Genomic DNA.
 DR EMBL; S67482; -; NOT ANNOTATED CDS; Genomic DNA.
 DR EMBL; X54876; CAA38657.1; ALT_SEQ; Genomic DNA.
 DR EMBL; M14423; AAA37333.1; -; mRNA.
 DR EMBL; M17491; AAA37334.1; -; Genomic DNA.
 DR EMBL; K03036; AAA37332.1; -; Genomic DNA.
 DR EMBL; K03029; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K03030; AAA37332.1; JOINED; Genomic DNA.

DR EMBL; K03031; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K03032; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K03033; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K03034; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K03035; AAA37332.1; JOINED; Genomic DNA.
 DR EMBL; K01688; AAA37330.1; -; Genomic DNA.
 DR EMBL; X06753; CAA29927.1; -; Genomic DNA.
 DR EMBL; X15896; CAA33904.1; -; Genomic DNA.
 DR EMBL; X57981; CAA41046.1; -; Genomic DNA.
 DR PIR; S57243; S21626.
 DR Ensembl; ENSMUSG00000001506; Mus musculus.
 DR MGI; MGI:88467; Colla1.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; WVFC; 1.
 DR PROSITE; PS0184; WVFC_2; 1.
 DR Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
 KW Pyridone carboxylic acid; Repeat; Signal; Structural protein.
 FT SIGNAL 1 22
 FT PROPEP 23 151 N-terminal propeptide.
 FT CHAIN 152 1207 Collagen alpha 1(I) chain.
 FT PROPEP 1208 1453 C-terminal propeptide.
 FT DOMAIN 29 87
 FT REGION 152 167 Nonhelical region (N-terminal).
 FT REGION 168 181 Triple-helical region.
 FT REGION 182 1207 Nonhelical region (C-terminal).
 FT MOTIF 734 736 Cell attachment site (potential).
 FT MOTIF 1082 1084 Cell attachment site (potential).
 FT MOD_RES 152 152 Pyridone carboxylic acid (By similarity).
 FT MOD_RES 160 160 Allysine (By similarity).
 FT MOD_RES 254 254 5-hydroxylysine (By similarity).
 FT MOD_RES 1153 1153 3-hydroxyproline (By similarity).
 FT CARBOHYD 56 56 N-linked (GlcNAc...) (potential).
 FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
 FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
 FT CONFLICT 81 81 G -> E (in Ref. 2 and 3).
 FT CONFLICT 106 106 G -> D (in Ref. 2 and 3).
 FT CONFLICT 1202 1202 D -> G (in Ref. 2).
 FT CONFLICT 1219 1219 A -> E (in Ref. 2).
 FT CONFLICT 1222 1222 T -> A (in Ref. 2).
 FT CONFLICT 1335 1335 T -> A (in Ref. 2).
 FT CONFLICT 1399 1400 RV -> TL (in Ref. 2).
 FT CONFLICT 1450 1450 A -> V (in Ref. 8).
 SQ SEQUENCE 1453 AA; 137943 MW; 3B802E535DF81808 CRC64;
 Query Match 90.5%; Score 38; DB 1; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAEGSPG 7
 |||||
 Db 1008 GAEGSPG 1014
 RESULT 21
 Q63079 RAT
 ID Q63079 RAT PRELIMINARY; PRT; 1453 AA.
 AC Q63079
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Collagen alpha1 (Fragment).
 OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Rattus.
 OK NCBI_TaxID=10116;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
 RX MEDLINE=99163824; PubMed=10065941;
 RA Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;
 RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
 formation in the rat.";
 RL J. Dent. Res. 78:11-19(1999).
 DR EMBL; Z78279; CAB01633.1; -; mRNA.
 DR Ensembl; ENSRNOG00000003897; Rattus norvegicus.
 DR GO; GO:0005581; C:cytoplasm; IEA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:000515; F:protein binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR00885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS50184; VWFC_2; 1.
 DR Collagen.
 KW NON TER
 FT
 SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;
 Query Match 90.5%; Score 38; DB 2; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GAEGSPG 7
 Db 1008 GAEGSPG 1014
 RESULT 22
 Q810J9 MOUSE PRELIMINARY; PRT; 1453 AA.
 ID Q810J9
 AC Q810J9
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Procollagen, type I, alpha 1.
 GN Name=Collal; ORFNames=RP23-112C19.9-001;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OK NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalski U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N; TISSUE=Colon;
 RA Director MGC Project; to the EMBL/GenBank/DBJ databases.
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Chapman J.;
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RA Smith M.;
 RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC050014; AAH50014.1; -; mRNA.
 DR EMBL; AL606480; CAI23970.1; -; Genomic DNA.
 DR EMBL; AL662790; CAI25880.1; -; Genomic DNA.
 DR EMBL; AL662790; CAI23970.1; JOINED; Genomic DNA.
 DR EMBL; AL606480; CAI25880.1; JOINED; Genomic DNA.
 DR Ensembl; ENSMUSG0000001506; Mus musculus.
 DR MGI; MGI:88467; Colla1.
 DR GO; GO:0005615; C:extracellular space; TAS.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR00885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS50184; VWFC_2; 1.
 DR Collagen.
 KW
 SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BB9A1D5EA CRC64;
 Query Match 90.5%; Score 38; DB 2; Length 1453;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GAEGSPG 7
 Db 1008 GAEGSPG 1014
 RESULT 23
 COL1A1 CANFA
 ID COL1A1 CANFA STANDARD; PRT; 1460 AA.
 AC Q9XSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Name=COL1A1;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Canis.
 OK NCBI_TaxID=9615;
 RN [1]
 RP NUCLEOTIDE SEQUENCE, AND VARIANT O1 ALA-208.
 RC TISSUE=Skin;

```

RX MEDLINE=21023337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of
RT canine osteogenesis imperfecta."
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -!- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
CC (OI).
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 WFPC domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AF153062; AAD34619.1; -; mRNA.
CC Ensembl; ENSCAPG0000017018; Canis familiaris.
CC InterPro; IPR008161; Clg_helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC InterPro; IPR001007; WVF_C.
CC Pfam; PF01410; COLFI; 1.
CC Pfam; PF01391; Collagen; 18.
CC ProDom; PD000007; Clg_helix; 2.
CC ProDom; PD002078; Fib_collagen_C; 1.
CC PROSITE; PS01208; WFPC_1; 1.
CC PROSITE; PS0184; WFPC_2; 1.
CC Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Pyrrolidone carboxylic acid; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 WFPC.
FT REGION 158 174 Nonhelical region (N-terminal).
FT REGION 175 1188 Triple-helical region.
FT REGION 1189 1214 Nonhelical region (C-terminal).
FT MOTIF 741 743 Cell attachment site (Potential).
FT MOTIF 1089 1091 Cell attachment site (Potential).
FT MOD_RES 158 158 Pyrrolidone carboxylic acid (By
FT similarity).
FT MOD_RES 166 166 Allylsine (By similarity).
FT MOD_RES 261 261 5-hydroxylysine (By similarity).
FT MOD_RES 1160 1160 3-hydroxyproline (By similarity).
FT CARBOHYD 261 261 O-linked (Gal... ) (By similarity).
FT CARBOHYD 1361 1361 N-linked (GlcNAc... ) (By similarity).
FT VARIANT 208 208 G -> A (in OI; severe).
FT SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 90.5%; Score 38; DB 1; Length 1460;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||||
Db 1015 GAEGSPG 1021

RESULT 24
O76045 HUMAN
ID O76045 HUMAN PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

```

```

DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=85130570; PubMed=2857713;
RA Chu M.-J., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=284343; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1)."
RL Gene 67:105-115(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen."
RL Biochem. J. 253:919-922(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins."
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98107942; PubMed=9443882;
RA Korhko J., Ala-Korhko L., De Paape A., Nuytinck L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I;
RT identification of common sequences of null-allele mutations."
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RA Korhko J.M., Earley J.J., Nuytinck L., DePaape A., Prockop D.J.,
RA Ala-Korhko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF017178; AAB94054.2; -; Genomic_DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; WVF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.

```

DR ProDom; PD002078; Fib collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWC_1; 1.
 DR PROSITE; PS50184; VWC_2; 1.
 KW Collagen.
 SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 90.5%; Score 38; DB 2; Length 1461;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
 |||||
 DB 1016 GAEGSPG 1022

RESULT 25

COL1A1 HUMAN STANDARD; PRT; 1464 AA.
 AC P02452; P78441; Q13896; Q13903; Q14037; Q14992; Q15176;
 AC Q15201; Q16050; Q7K230; Q7K233; Q8IV15; Q9UML6; Q9UMW7;
 DT 21-JUL-1986 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Name=COL1A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE, AND VARIANT SER-1434.
 RA Daigleish R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 1-589.
 RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
 RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
 RA Pretorius P.J.;
 RT "Complete nucleotide sequence of the region encompassing the first
 RT twenty-five exons of the human pro alpha 1(I) collagen gene
 RT (COL1A1).";
 RL Gene 67:105-115(1988).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 1-472.
 RX MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
 RA Jaenisch R., Prockup D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
 RT of human type I procollagen.";
 RL Biochem. J. 253:919-922(1988).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 1-181.
 RX MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 RT conservation of a pattern of introns and exons.";
 RL Nature 310:337-340(1984).
 RN [5]
 RP PROTEIN SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
 RN TISSUE=Skin;
 RC Click E.M., Bornstein P.;
 RX MEDLINE=71038625; PubMed=5529814;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 RT the alpha 1 and alpha 2 chains of human skin collagen.";
 RL Biochemistry 9:4699-4706(1970).
 RN [6]
 RP NUCLEOTIDE SEQUENCE OF 425-1464.
 RX MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,

RA Prockup D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
 RT of structures that are conserved during evolution.";
 RL Biochemistry 22:5213-5223(1983).
 RN [7]
 RP NUCLEOTIDE SEQUENCE OF 472-607.
 RX MEDLINE=85104934; PubMed=2981843;
 RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
 RT "Multixon deletion in an osteogenesis imperfecta variant with
 RT increased type III collagen mRNA.";
 RL J. Biol. Chem. 260:691-694(1985).
 RN [8]
 RP NUCLEOTIDE SEQUENCE OF 488-625.
 RX MEDLINE=85190598; PubMed=3857621;
 RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
 RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
 RT I collagen chain in a lethal form of osteogenesis imperfecta.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
 RN [9]
 RP NUCLEOTIDE SEQUENCE OF 1179-1464, VARIANTS OI-II HIS-1277; ARG-1388
 RP AND 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
 RX MEDLINE=93352646; PubMed=8349697;
 RA Chessler S.D., Wallis G.A., Byers P.H.;
 RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
 RT chain of type I collagen result in defective chain association and
 RT produce lethal osteogenesis imperfecta.";
 RL J. Biol. Chem. 268:18218-18225(1993).
 RN [10]
 RP NUCLEOTIDE SEQUENCE OF 1229-1454.
 RC TISSUE=Bone;
 RX MEDLINE=88124208; PubMed=3340531;
 RA Maekela J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 RT domain.";
 RL Nucleic Acids Res. 16:349-349(1988).
 RN [11]
 RP NUCLEOTIDE SEQUENCE OF 1-34.
 RX MEDLINE=85130970; PubMed=2857713;
 RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
 RT "Fine structural analysis of the human pro-alpha 1(I) collagen gene.
 RT Promoter structure, AluI repeats, and polymorphic transcripts.";
 RL J. Biol. Chem. 260:2315-2320(1985).
 RN [12]
 RP NUCLEOTIDE SEQUENCE OF 1-34.
 RX MEDLINE=88097389; PubMed=3480516;
 RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
 RT "Regulatory elements in the first intron contribute to transcriptional
 RT control of the human alpha 1(I) collagen gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
 RN [13]
 RP NUCLEOTIDE SEQUENCE OF 1-44.
 RX MEDLINE=88033098; PubMed=2822714;
 RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
 RA de Wet W.;
 RT "DNA sequences in the first intron of the human pro-alpha 1(I)
 RT collagen gene enhance transcription.";
 RL J. Biol. Chem. 262:15151-15157(1987).
 RN [14]
 RP PROTEIN SEQUENCE OF 33-52.
 RX MEDLINE=90202908; PubMed=2318855;
 RA Witz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
 RA Rao V.H., Hollister D.W.;
 RT "In vivo and in vitro noncovalent association of excised alpha 1(I)
 RT amino-terminal propeptides with mutant pro alpha 2(I) collagen chains
 RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
 RT VII.";
 RL J. Biol. Chem. 265:6312-6317(1990).
 RN [15]
 RP NUCLEOTIDE SEQUENCE OF 156-183.
 RX MEDLINE=89356643; PubMed=2767050;
 RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
 RA Bateman J.F.;

RT "A base substitution in the exon of a collagen gene causes alternative
RT splicing and generates a structurally abnormal polypeptide in a
RT patient with Ehlers-danlos syndrome type VII.";
RL EMBO J. 8:1705-1710(1989).

[16]

RN PROTEIN SEQUENCE OF 175-187 AND 274-289.

RX MEDLINE=90382436; PubMed=2169412;
RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
RA Muller P.K.;

RT "A critical crosslink region in human-bone-derived collagen type I.
RT Specific cleavage site at residue Leu95.";
RL Eur. J. Biochem. 192:153-159(1990).

[17]

RN PROTEIN SEQUENCE OF 263-268.

RC TISSUE=Skin;

RX MEDLINE=71001508; PubMed=4319110;

RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;

RT "A comparative study of glycopeptides derived from selected vertebrate
RT collagens. A possible role of the carbohydrate in fibril formation.";
RL J. Biol. Chem. 245:5042-5048(1970).

[18]

RN NUCLEOTIDE SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045
RP AND 1143-1162.

RX MEDLINE=90326017; PubMed=2374517;

RA Labhard M.E., Hollister D.W.;

RT "Segmental amplification of the entire helical and telopeptide regions
RT of the cDNA for human alpha 1 (I) collagen.";
RL Matrix 10:124-130(1990).

[19]

RN NUCLEOTIDE SEQUENCE OF 425-490; 965-1024; 999-1039 AND 1453-1464.

RX MEDLINE=83064528; PubMed=6183642;

RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;

RT "Cloning and characterization of five overlapping cDNAs specific for
RT the human pro alpha 1(I) collagen chain.";
RL Nucleic Acids Res. 10:5925-5934(1982).

[20]

RN NUCLEOTIDE SEQUENCE OF 710-745, AND VARIANT OI-II ARG-728.

RX MEDLINE=90252792; PubMed=2339700;

RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;

RT "Variable expression of osteogenesis imperfecta in a nuclear family is
RT explained by somatic mosaicism for a lethal point mutation in the
RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
RL Am. J. Hum. Genet. 46:1034-1040(1990).

[21]

RN NUCLEOTIDE SEQUENCE OF 746-781, AND VARIANT OI-III SER-767.

RX MEDLINE=95187161; PubMed=7881420;

RA Forlino A., Zolazzi F., Valli M., Pignatti P.F., Cetta G.,

RA Brunelli P.C., Mottes M.;

RT "Severe (type III) osteogenesis imperfecta due to glycine
RT substitutions in the central domain of the collagen triple helix.";
RL Hum. Mol. Genet. 3:2201-2206(1994).

[22]

RN NUCLEOTIDE SEQUENCE OF 1187-1220, AND VARIANT CVS-1195.

RX MEDLINE=89008319; PubMed=3170557;

RA Cohn D.H., Apone S., Byre D.R., Starman B.J., Andreassen P.,

RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;

RT "Substitution of cysteine for glycine within the carboxyl-terminal

RT telopeptide of the alpha 1 chain of type I collagen produces mild

RT osteogenesis imperfecta.";

RL J. Biol. Chem. 263:14605-14607(1988).

[23]

RN NUCLEOTIDE SEQUENCE OF 1440-1464.

RX MEDLINE=90110490; PubMed=2295701;

RA Willing M.C., Cohn D.H., Byers P.H.;

RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I

RT collagen predicts an elongated pro alpha 1(I) chain and results in

RT osteogenesis imperfecta type I.";

RL J. Clin. Invest. 85:282-290(1990).

[24]

RN NUCLEOTIDE SEQUENCE OF 1454-1464.

RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;

RA Maatta A., Bornstein P., Penttinen R.P.;

RT "highly conserved sequences in the 3'-untranslated region of the

Query Match 90.5%; Score 38; DB 1; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7

DB 1019 GAEGSPG 1025

|||||

RESULT 26

Q8N473 HUMAN

ID Q8N473 HUMAN PRELIMINARY; PRT; 1464 AA.

AC Q8N473;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Alpha 1 type I collagen, preproprotein.

GN Name=COL1A1;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

OC Homo.

OX NCBI_TaxID=9606;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Brain;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Brain;

RG NIH MGC Project;

RL Submitted (Aug-2002) to the EMBL/GenBank/DBSJ databases.

DR EMBL; BC036531; AAH36531.1; -; mRNA.

DR Ensembl; ENSG00000108821; Homo sapiens.

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005737; C:cyclopasm; IEA.

DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.

DR GO; GO:0008817; P:phosphate transport; IEA.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib_collagen_C.

DR InterPro; IPR001007; VWF_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR ProDom; PD000007; Clg_helix; 3.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.

DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01208; VWF_C_1; 1.

DR PROSITE; PS0184; VWFC_2; 1.

KW Collagen.

SQ SEQUENCE 1464 AA; 139011 MW; B0581F8D1C89DD88 CRC64;

Query Match 90.5%; Score 38; DB 2; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
|||||
Db 1019 GAEGSPG 1025

RESULT 27
Q59F64 HUMAN PRELIMINARY; PRT; 1467 AA.
AC Q59F64
DT 10-MAY-2005 (TRENBLrel. 30, Created)
DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Collagen alpha 1 chain variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE-Spleen;
RA Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,
Ohara O., Nagase T., Kikuno F.R.;
RT "None Title";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB209597; BAD92834.1; -; mRNA.
KW Collagen.
FT NON_TER 1
SQ SEQUENCE 1467 AA; 139196 MW; 976847FDE93E7945 CRC64;
Query Match 90.5%; Score 38; DB 2; Length 1467;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
|||||
Db 1022 GAEGSPG 1028

RESULT 28
Q597P9 CANFA PRELIMINARY; PRT; 1688 AA.
AC Q597P9
DT 10-MAY-2005 (TRENBLrel. 30, Created)
DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Collagen type IV alpha 4 chain.
GN Name=COL4A4;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OX NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Wiersma A.C., Millon L.V., Van Dongen A.M., Van Oost B.A.,
Bannasch D.L.;
RT "Canine COL4A3 and COL4A4: cDNA sequencing, mapping and
characterization as candidates for familial renal disease in the
Norwegian Elkhound";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY263363; AAP88582.1; -; mRNA.
KW Collagen.
SQ SEQUENCE 1688 AA; 164573 MW; D26F4B6F6AF0BFOA CRC64;
Query Match 90.5%; Score 38; DB 2; Length 1688;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPG 7
|||||
Db 530 GAEGSPG 536

RESULT 29
Q7VVJ2 BORPE PRELIMINARY; PRT; 2553 AA.
AC Q7VVJ2
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Adhesin.
GN Name=fhas; OrderedLocusNames=BP2667;
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Tohama 1 / ATCC BAA-589 / NCTC 13251;
RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ngi1227;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.R.,
Harris D.B., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsels K.,
Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
Bordetella parapertussis and Bordetella bronchiseptica";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640419; CAB42943.1; -; Genomic_DNA.
DR InterPro; IPR008619; Fil_haemagg
DR InterPro; IPR010069; Fil_hemag_20aa.
DR InterPro; IPR008638; Haemagg_act_N.
DR Pfam; PF05594; Fil_haemagg; 8.
DR Pfam; PF05860; Haemagg_act; 1.
DR TIGRFAMs; TIGR01901; adhes_NPXG; 1.
DR TIGRFAMs; TIGR01731; fil_hemag_20aa; 4.
KW Complete proteome.
SQ SEQUENCE 2553 AA; 263697 MW; 400F779E0AD581DF CRC64;

Query Match 90.5%; Score 38; DB 2; Length 2553;
Best Local Similarity 87.5%; Pred. No. 9.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|||||
Db 2090 GAEGSPGL 2097

RESULT 30
Q5AQ38 EMENI PRELIMINARY; PRT; 435 AA.
AC Q5AQ38
DT 10-MAY-2005 (TRENBLrel. 30, Created)
DT 10-MAY-2005 (TRENBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TRENBLrel. 30, Last annotation update)
DE Predicted protein.
GN ORFNames=AN9352.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;

RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Baatien V., Bloom T., Boguslavsky J.,
RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Chospel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gierre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meidrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nieleen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Teefave S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACD01000172; EAA66419.1; -; Genomic DNA.
SQ SEQUENCE 435 AA; 48158 MW; 526C008F2D3A7528 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 435;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGGSPGL 8
|| |||||
DB 118 GAGGSPGL 125

RESULT 31
Q06452 EPHMU
ID Q06452 EPHMU PRELIMINARY; PRT; 812 AA.
AC Q06452;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Eml1 alpha (Fragment).
GN Names=COLF1;
OS Ephydatia muelleri (Mueller's freshwater sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Haplosclerida; Spongiillidae; Ephydatia.
OX NCBI_TaxID=6052;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=94047120; PubMed=8230249; DOI=10.1007/BF00175502;
RA Exposito J.Y., van der Rest M., Garrone R.;
RT "The complete intron/exon structure of Ephydatia muelleri fibrillar
RT collagen gene suggests a mechanism for the evolution of an ancestral
RT gene module.";
RL J. Mol. Evol. 37:254-259 (1993).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Exposito J.;
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X69818; CAA49472.1; -; Genomic_DNA.
DR PIR; S31521; S31521.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 12.
KW Collagen.
FT NON TER 812 812
SQ SEQUENCE 812 AA; 72280 MW; 326573F37E46D50F CRC64;

Query Match 88.1%; Score 37; DB 2; Length 812;
Best Local Similarity 87.5%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGGSPGL 8
|| |||||
DB 84 GAGGSPGL 91

RESULT 32
Q8CFM4 MOUSE
ID Q8CFM4 MOUSE PRELIMINARY; PRT; 998 AA.
AC Q8CFM4;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Col3a1 protein (Fragment).
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013626; AAH13626.1; -; mRNA.
DR Ensembl; ENSMUSG00000026043; Mus musculus.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 12.
DR Prodom; PD000007; Clg_helix; 2.
KW Collagen.
FT NON TER 1 1
SQ SEQUENCE 998 AA; 95451 MW; 2AB4BA953B7084A6 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 998;
Best Local Similarity 87.5%; Pred. No. 5.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGGSPGL 8


```

FT MOD RES 792 4-hydroxyproline.
FT MOD RES 810 5-hydroxylysine (Probable).
FT MOD RES 815 3-hydroxyproline (partial).
FT MOD RES 816 4-hydroxyproline.
FT MOD RES 843 4-hydroxyproline.
FT MOD RES 849 4-hydroxyproline.
FT MOD RES 855 4-hydroxyproline.
FT MOD RES 861 4-hydroxyproline.
FT MOD RES 867 4-hydroxyproline.
FT MOD RES 888 4-hydroxyproline.
FT MOD RES 894 4-hydroxyproline.
FT MOD RES 903 4-hydroxyproline.
FT MOD RES 915 4-hydroxyproline.
FT MOD RES 927 5-hydroxylysine (Probable).
FT MOD RES 933 5-hydroxylysine (partial).
FT MOD RES 936 5-hydroxylysine (Probable).
FT MOD RES 939 5-hydroxylysine.
FT MOD RES 945 4-hydroxyproline.
FT MOD RES 954 4-hydroxyproline (partial).
FT MOD RES 963 4-hydroxyproline.
FT MOD RES 966 4-hydroxyproline.
FT MOD RES 984 4-hydroxyproline.
FT MOD RES 990 4-hydroxyproline.
FT MOD RES 1010 3-hydroxyproline (partial).
FT MOD RES 1011 4-hydroxyproline.
FT MOD RES 1013 3-hydroxyproline (partial).
FT MOD RES 1014 4-hydroxyproline.
FT MOD RES 1016 3-hydroxyproline.
FT MOD RES 1017 3-hydroxyproline.
FT MOD RES 1019 3-hydroxyproline.
FT MOD RES 1020 4-hydroxyproline.
FT CARBOHYD 96 O-linked (Gal. . .) (Probable).
FT CARBOHYD 108 O-linked (Gal. . .) (Probable).
FT CARBOHYD 192 O-linked (Gal. . .) (Probable).
FT CARBOHYD 261 O-linked (Gal. . .) (Probable).
FT CARBOHYD 279 O-linked (Gal. . .) (Probable).
FT CARBOHYD 573 O-linked (Gal. . .) (Probable).
FT CARBOHYD 612 O-linked (Gal. . .) (Probable).
FT CARBOHYD 657 O-linked (Gal. . .) (Probable).
FT CARBOHYD 738 O-linked (Gal. . .) (Probable).
FT CARBOHYD 765 O-linked (Gal. . .) (Probable).
FT CARBOHYD 810 O-linked (Gal. . .) (Probable).
FT CARBOHYD 927 O-linked (Gal. . .) (Probable).
FT CARBOHYD 936 O-linked (Gal. . .) (Probable).
FT VARIANT 936 P -> A.
FT UNSURE 96 96
FT UNSURE 108 108
FT UNSURE 192 192
FT UNSURE 261 261
FT UNSURE 279 279
FT UNSURE 573 573
FT UNSURE 612 612
FT UNSURE 657 657
FT UNSURE 738 738
FT UNSURE 765 765
FT UNSURE 810 810
FT UNSURE 927 927
FT UNSURE 936 936
FT NON_TER 1 1
FT 1027 1027

```

Query Match 88.1%; Score 37; DB 1; Length 1027;
 Best Local Similarity 87.5%; Pred. No. 5.6e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 Db 739 GARGSPGL 746

RESULT 34
 Q8K173 MOUSE
 ID Q8K173_MOUSE PRELIMINARY; PRT; 1222 AA.

```

AC Q8K173;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Col3a1 protein (Fragment).
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RL "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strausberg R.;
RA Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC028248; AAB28248.1; -; mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 16.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
FT NON_TER 1 1
FT SEQUENCE 1222 AA; 115140 MW; A409CA00D82765E4 CRC64;
SQ SEQUENCE 1222 AA; 115140 MW; A409CA00D82765E4 CRC64;

```

Query Match 88.1%; Score 37; DB 2; Length 1222;
 Best Local Similarity 87.5%; Pred. No. 6.7e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
 |||||
 Db 474 GASGSPGL 481

RESULT 35
 Q53RW9 HUMAN
 ID Q53RW9_HUMAN PRELIMINARY; PRT; 1324 AA.
 AC Q53RW9;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Hypothetical protein COL4A4 (Fragment).
 GN Name=COL4A4;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Maupin R., Spalding L., Mclellan M., Haglund K.;
 RT "The sequence of Homo sapiens BAC clone RP11-495O23.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA Waterston R.H.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Waterston R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RA Wilson R.K.;
 RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC079235; AAY14670.1; -; Genomic_DNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 KW Collagen; Hypothetical protein.
 FT NON_TER 1324 1324
 SQ SEQUENCE 1324 AA; 125826 MW; 556BF2BB635090F3 CRC64;

 Query Match 88.1%; Score 37; DB 2; Length 1324;
 Best Local Similarity 87.5%; Pred. No. 7.2e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 Qy 1 GAEGSPGL 8
 Db 533 GAEGPPGL 540

 RESULT 36
 Q26634_STRPU PRELIMINARY; PRT; 1414 AA.
 AC Q26634;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Alpha-1 collagen.
 GN Name=COL1A1alpha;
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
 OC Strongylocentrotus.
 OX NCBI_TaxID=7668;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=92348411; PubMed=1639795;
 RA Exposito J.-Y., D'Alessio M., Solursh M., Ramirez F.;
 RT "Sea urchin collagen evolutionarily homologous to vertebrate Pro-alpha-2(I) collagen.";
 RL J. Biol. Chem. 267:15559-15562(1992).
 DR EMBL; M92040; AAA30035.1; -; mRNA.
 DR PIR; S23809; S23809
 DR GO; GO:0005581; C:collagen; IEA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.
 DR SMART; SM00038; COLFI; 1.
 KW Collagen. 1414 AA; 133024 MW; 94D9CDA71A9FD73D CRC64;
 SQ SEQUENCE 1414 AA; 133024 MW; 94D9CDA71A9FD73D CRC64;

 Query Match 88.1%; Score 37; DB 2; Length 1414;
 Best Local Similarity 87.5%; Pred. No. 7.8e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 Qy 1 GAEGSPGL 8
 Db 324 GARGSPGL 331

 RESULT 37
 CO3A1_MOUSE
 ID CO3A1_MOUSE STANDARD; PRT; 1464 AA.
 AC P08121; O61429; Q9CRN7;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN Name=Col3a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
 RX MEDLINE=95011609; PubMed=7926795; DOI=10.1016/0378-1119(94)90061-2;
 RA Toman D., de Crombrughe B.;
 RT "The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA sequence.";
 RL Gene 147:161-168(1994).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 1-488.
 RX MEDLINE=88167858; PubMed=3443309; DOI=10.1016/0378-1119(87)90117-X;
 RA Wood L., Theriault N., Vogel G.;
 RT "Complete nucleotide sequence of the N-terminal domains of the murine alpha-1 type-III collagen chain.";
 RL Gene 61:225-230(1987).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 1-28.
 RX MEDLINE=85131189; PubMed=3972847;
 RA Liau G., Mudryj M., de Crombrughe B.;
 RT "Identification of the promoter and first exon of the mouse alpha 1 (III) collagen gene.";

J. Biol. Chem. 260:3773-3777(1985).

[5] NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 810-1464.

RC STRAIN=C57BL/6J; TISSUE=Embryonic head; DOI=10.1038/nature01266;

RA MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;

RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,

RA Nikaide I., Oseato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,

RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,

RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,

RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,

RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,

RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Fough J.K.,

RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,

RA Hammond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,

RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,

RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,

RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,

RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,

RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,

RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,

RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,

RA Verardo R., Wagner L., Wawlesed C., Wang Y., Watanabe Y., Wells C.,

RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,

RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,

RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,

RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,

RA Hara A., Hashizume W., Imocani K., Ienil Y., Itoh M., Kagawa I.,

RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,

RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,

RA Birney E., Hayashizaki Y.,

RT "Analysis of the mouse transcriptome based on functional annotation of

RT 60,770 full-length cDNAs.";

RL Nature 420:563-573(2002).

[6] NUCLEOTIDE SEQUENCE OF 1442-1464.

RC STRAIN=C57BL/6J;

RC MEDLINE=91274355; PubMed=2054384;

RA Metaxeranta M., Toman D., de Crombrughe B., Vuorio E.;

RT "Specific hybridization probes for mouse type I, II, III and IX

RT collagen mRNAs.";

RL Biochim. Biophys. Acta 1089:241-243(1991).

CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues

CC along with type I collagen.

CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are

CC linked to each other by interchain disulfide bonds. Trimers are

CC also cross-linked via hydroxylysines.

CC -!- PTM: Proline residues at the third position of the tripeptide

CC repeating unit (G-X-Y) are hydroxylated in some or all of the

CC chains.

CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to

CC the oxygen atom of a post-translationally added hydroxyl group (By

CC similarity).

CC -!- SIMILARITY: Belongs to the fibrillar collagen family.

CC -!- SIMILARITY: Contains 1 WFCC domain.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not

CC removed.

CC -----

CC EMBL; X52046; CAA36279.1; -; Genomic DNA.

DR EMBL; BC043089; AAH43089.1; -; mRNA.

DR EMBL; BC058724; AAH58724.1; -; mRNA.

DR EMBL; M18933; AAA37338.1; -; mRNA.

DR EMBL; K03037; -; NOT_ANNOTATED_CDS; Genomic DNA.

DR EMBL; AK019448; BAB31724.1; -; mRNA.

DR EMBL; X57983; CAA41048.1; -; Genomic DNA.

DR F01; A27353; A27353.

DR F01; X59856; X59856.

DR EMBL; ENSMUSG0000001506; Mus musculus.

DR MGI; MGI:88453; Col3a1.

DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib_collagen_C.

DR InterPro; IPR001007; WF_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 17.

DR ProDom; PD000007; C1g_helix; 3.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR PROSITE; PS01208; WF_C; 1.

DR PROSITE; PS0184; WF_C; 2.

KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat;

KW SIGNAL; Structural protein.

FT SIGNAL 1 23 By similarity.

FT PROPEP 24 154 N-terminal propeptide.

FT CHAIN 155 1203 Collagen alpha 1(III) chain.

FT PROPEP 1204 1464 C-terminal propeptide.

FT DOMAIN 31 90

FT REGION 155 169 Nonhelical region (N-terminal).

FT REGION 170 1195 Triple-helical region.

FT REGION 1196 1464 Nonhelical region (C-terminal).

FT MOD_RES 262 262 5-hydroxylysine (By similarity).

FT MOD_RES 283 283 5-hydroxylysine (By similarity).

FT MOD_RES 859 859 5-hydroxylysine (By similarity).

FT MOD_RES 976 976 5-hydroxylysine (By similarity).

FT MOD_RES 1093 1093 5-hydroxylysine (By similarity).

FT MOD_RES 1105 1105 5-hydroxylysine (By similarity).

FT CARBOHYD 262 262 O-linked (Gal..) (By similarity).

FT DISULFID 1195 1195 Interchain (By similarity).

FT DISULFID 1196 1196 Interchain (By similarity).

FT SEQUENCE 1464 AA; 138943 MW; 2104EC27A886090B CRC64;

SQ

Query Match 88.1%; Score 37; DB 1; Length 1464;

Best Local Similarity 87.5%; Pred. No. 8e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGSGPGL 8

Db 716 GAGSGPGL 723

RESULT 38

Q8BK2_MOUSE

ID Q8BK2_MOUSE PRELIMINARY; PRT; 1464 AA.

AC Q8BK2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched

DE library, clone:CI30072F01 product:COLLAGEN ALPHA 1(III) CHAIN, full

DE insert sequence.

GN Name=Col3a1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

[1]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Head;

RC MEDLINE=9279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P., Hayashizaki Y.;

RA "High-efficiency full-length cDNA cloning.";

RT Meth. Enzymol. 303:19-44(1999).

RL [2]

RN NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Head;

RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaïdo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaïdo M.F.,
RA Brownstein S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Gustincich S., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuina T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yanamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohmoto N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tonaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK048546; BAC33370.1; -, mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; C1g_helix; 3.

DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138946 MW; 1E4ED9539EF42C12 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 1464;
Best Local Similarity 87.5%; Pred. No. 8e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 1;

Qy 1 GAEGSPGL 8
||| |||||
Db 716 GASGSPGL 723

RESULT 39
Q8BLW4_MOUSE
AC Q8BLW4_MOUSE PRELIMINARY; PRT; 1464 AA.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male aorta and vein cDNA, RIKEN full-length
DE enriched library, clone:A530083N24 product:COLLAGEN ALPHA 1(III)
DE CHAIN, full insert sequence.
DE Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaïdo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Boffelli D., Bojunga N., Carninci P., de Bonaïdo M.F.,
RA Blake J., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein S., Hill D., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]

RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RT Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaki S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK041115; BAC30826.1; -; mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138938 MW; 91F3246D90818449 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 1464;
Best Local Similarity 87.5%; Pred. No. 8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|||
Db 716 GASGSPGL 723

RESULT 40
Q7TT32 MOUSE
ID Q7TT32_MOUSE PRELIMINARY; PRT; 1464 AA.
AC Q7TT32;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Procollagen, type III, alpha 1.
GN Name=Coll3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
[1]
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=2238625; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.C., Grimwood J., Schmutz J., Myers R.M.,
RA Rodriguez A.C., Touchman J.W., Green E.D., Dickson M.C.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.W., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052398; AAH52398.1; -; mRNA.
DR Ensembl; ENSMUSG0000001506; Mus musculus.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;

Query Match 88.1%; Score 37; DB 2; Length 1464;
Best Local Similarity 87.5%; Pred. No. 8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|||
Db 716 GASGSPGL 723

RESULT 41
Q5DTG2 MOUSE
ID Q5DTG2_MOUSE PRELIMINARY; PRT; 1467 AA.
AC Q5DTG2;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)

10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
MKTAA4231 protein (Fragment)
Name=COL3a1; Synonyms=MKTAA4231;
Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Fetal brain;
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Nagase T., Ohara O.,
RA Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene.
RT The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs
RT Identified by Screening of Terminal sequences of cDNA Clones Randomly
RT Sampled from Size-Fractionated Libraries.";
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK220558; BAD90325.1; -, mRNA.
DR MGI: MGI-88453; Col3a1.
DR GO: GO:0005581; C:collagen; IEA.
DR GO: GO:0005737; C:cytoplasm; IEA.
DR GO: GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO: GO:0006817; P:phosphate transport; IEA.
DR InterPro: IPR008161; Clq helix.
DR InterPro: IPR008160; Collagen.
DR InterPro: IPR000885; Fib-collagen_C.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF01410; COLFI; 1.
DR Pfam: PF01391; Collagen; 17.
DR ProDom: PD000007; Clq helix; 3.
DR ProDom: PD002078; Fib-collagen_C; 1.
DR SMART: SM00038; COLFI; 1.
DR SMART: SM00214; VWF; 1.
DR PROSITE: PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE: PS50184; VWF_C_2; 1.
KW Collagen.
FT NON TER.
SQ SEQUENCE 1467 AA; 139335 MW; 73274A578803DCF8 CRC64;
Query Match 88.1%; Score 37; DB 2; Length 1467;
Best Local Similarity 87.5%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAEGSPGL 8
Db 719 GASGSPGL 726
RESULT 42
CO4A4 HUMAN
ID CO4A4 HUMAN STANDARD; PRT; 1690 AA.
AC P53420;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 4(IV) chain precursor.
GN Name=COL4A4;
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Kidney;
RA Letnonen A., Mariyama M., Mochizuki T., Tryggvason K., Reeders S.T.;
RT "Complete primary structure of the human type IV collagen alpha 4(IV)
RT chain. Comparison with structure and expression of the other alpha
RT (IV) chains.";
RL J. Biol. Chem. 269:26172-26177(1994).

RN [2]
RP NUCLEOTIDE SEQUENCE OF 1-23.
RX MEDLINE=98196854; PubMed=9537506; DOI=10.1016/S0014-5793(98)00128-8;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Ninomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE OF 1219-1690.
RC TISSUE=Eye;
RX MEDLINE=933374047; PubMed=8365481; DOI=10.1016/0014-5793(93)80256-T;
RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
RT collagen chain.";
RL FEBS Lett. 330:122-128(1993).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1407-1507.
RX MEDLINE=93054733; PubMed=1429714;
RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha
RT 4 chain of basement membrane collagen type IV and assignment of the
RT gene to the distal long arm of human chromosome 2.";
RL J. Biol. Chem. 267:23753-23758(1992).
RN [5]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RX DOI=10.1002/(SICI)1098-1004(1997)9:6<477::AID-HUMU1>3.3.CO;2-H;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
RN [6]
RP VARIANT AS SER-1201.
RX MEDLINE=95078927; PubMed=7987396;
RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
RA Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
RA Smeets H.J.M., Reeders S.T.;
RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
RT collagen genes in autosomal recessive Alport syndrome.";
RL Nat. Genet. 8:77-82(1994).
RN [7]
RP VARIANT FBH GLU-897.
RX MEDLINE=96379660; PubMed=8787673;
RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
RT "Benign familial hematuria due to mutation of the type IV collagen
RT alpha4 gene.";
RL J. Clin. Invest. 98:1114-1118(1996).
RN [8]
RP VARIANTS AS, AND VARIANTS.
RX MEDLINE=99011253; PubMed=9792860;
RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
RA Cochot P., Gruenfeld J.-P., Falcoux J.-B., Gubler M.-C., Antignac C.;
RT "Determination of the genomic structure of the COL4A4 gene and of
RT novel mutations causing autosomal recessive Alport syndrome.";
RL Am. J. Hum. Genet. 63:1329-1340(1998).
CC -1- FUNCTION: type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and
CC entactin/nidogen.
CC -1- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -1- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -1- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -1- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.

CC -!- PTM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

CC -!- DISEASE: Defects in COL4A4 are a cause of autosomal recessive Alport syndrome (AS) [MIM:203780]. AS is an hereditary disorder characterized by progressive glomerulonephritis, renal failure, hematuria, ocular abnormalities and deafness. The recessive form occurs equally between males and females.

CC -!- DISEASE: Defects in COL4A4 are a cause of familial benign hematuria (FBH) [MIM:141200]; also known as thin basement membrane disease. FBH is characterized by persistent hematuria, an electron microscopically detectable thin glomerular basement membrane (GBM) and an autosomal dominant mode of inheritance. Renal function remains normal. In children, differentiation between FBH and AS can be difficult, because both disorders are manifested by persistent hematuria and thin GBM at that age.

CC -!- SIMILARITY: Belongs to the type IV collagen family.

CC -!- SIMILARITY: Contains 2 COL4C (collagen IV C-terminal) domains.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC -----

CC EMBL; X81053; CAA56943.1; -, mRNA.

CC EMBL; AB008496; BAA25065.1; -, Genomic_DNA.

CC EMBL; D17391; BAA04214.1; -, mRNA.

CC EMBL; A55360; CGHULB.

CC HSP; F08572; LL1.

CC HGNC; HGNC:2206; COL4A4.

CC MIM; 120131; -.

CC MIM; 141200; -.

CC MIM; 203780; -.

CC GO; GO:0005587; C:collagen type IV; IDA.

CC GO; GO:0005201; P:extracellular matrix structural constituent; TAS.

CC GO; GO:0042062; P:long-term strengthening of neuromuscular ju. . .; IDA.

CC InterPro; IPR008161; Clg helix.

CC InterPro; IPR008160; Collagen.

CC InterPro; IPR001442; Procollagn4_C.

CC Pfam; PF01413; C4; 2.

CC Pfam; PF01391; Collagen; 20.

CC ProDom; PD0000007; Clg helix; 9.

CC ProDom; PD003923; Procollagn4_C; 2.

CC SMART; SM00111; C4; 2.

CC Alport syndrome; Basement membrane; Collagen; Disease mutation;

CC Extracellular matrix; Glycoprotein; Hydroxylation; Polymorphism;

KW Repeat; Signal; Structural protein.

FT SIGNAL 1 38 Potential.

FT CHAIN 39 1690 Collagen alpha 4(IV) chain.

FT DOMAIN 1465 1572 COL4C 1.

FT DOMAIN 1573 1690 COL4C 2.

FT REGION 39 64 7S domain.

FT REGION 65 1459 Triple-helical region.

FT MOTIF 94 96 Cell attachment site (Potential).

FT MOTIF 145 147 Cell attachment site (Potential).

FT MOTIF 189 191 Cell attachment site (Potential).

FT MOTIF 310 312 Cell attachment site (Potential).

FT MOTIF 724 726 Cell attachment site (Potential).

FT MOTIF 785 787 Cell attachment site (Potential).

FT MOTIF 989 991 Cell attachment site (Potential).

FT MOTIF 1212 1214 Cell attachment site (Potential).

FT MOTIF 1206 1207 Cleavage (by collagenase) (By similarity).

FT SITE 142 142 N-linked (GlcNAc. . .) (Potential).

FT CARBOHYD 669 669 N-linked (GlcNAc. . .) (Potential).

FT DISULFID 1480 1569 Or C-1480 with C-1566 (By similarity).

FT DISULFID 1513 1566 Or C-1513 with C-1569 (By similarity).

FT DISULFID 1525 1531 By similarity.

FT DISULFID 1588 1686 Or C-1588 with C-1683 (By similarity).

FT DISULFID 1622 1683 Or C-1622 with C-1686 (By similarity).

FT DISULFID 1634 1641 By similarity.

FT VARIANT 441 446 Missing (in AS).

FT VARIANT 482 482 /FTid=VAR_008148.

FT VARIANT 545 545 P -> S (in dbSNP:3736633).

FT VARIANT 570 570 /FTid=VAR_022069.

FT VARIANT 897 897 G -> A (in dbSNP:1800516).

FT VARIANT 931 931 E -> Q.

FT VARIANT 1004 1004 /FTid=VAR_008150.

FT VARIANT 1030 1030 G -> E (in FBH).

FT VARIANT 1201 1201 /FTid=VAR_001912.

FT VARIANT 1402 1402 A -> T.

FT VARIANT 1572 1572 L -> P (in dbSNP:1800517).

FT CONFLICT 1659 1660 G -> V (in AS).

FT CONFLICT 1659 1660 /FTid=VAR_008153.

FT CONFLICT 1659 1660 G -> S (in AS).

FT CONFLICT 1659 1660 /FTid=VAR_001913.

FT CONFLICT 1659 1660 P -> S.

FT CONFLICT 1659 1660 /FTid=VAR_008154.

FT CONFLICT 1659 1660 P -> L (in AS).

FT CONFLICT 1659 1660 /FTid=VAR_008155.

FT CONFLICT 1659 1660 LQ -> FE (in Ref. 3).

SQ SEQUENCE 1690 AA; 164096 MW; 51E72F283A72BAAE CRC64;

Query Match 88.1%; Score 37; DB 1; Length 1690;

Best Local Similarity 87.5%; Pred No. 9.3e+02; Gaps 0;

Matches 7; Conservative 0; Mismatches 1; Indels 0;

QY 1 GAEGSPGL 8

DB 533 GAEGPPGL 540

RESULT 43

Q53WR1_HUMAN PRELIMINARY; PRT; 1690 AA.

AC Q53WR1_13-SEP-2005 (TRENBLrel. 31, Created)

DT 13-SEP-2005 (TRENBLrel. 31, Last sequence update)

DE 13-SEP-2005 (TRENBLrel. 31, Last annotation update)

DE Collagen alpha 4 type IV.

GN Name=COL4A4;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

OC Homo.

OX NCBI_TaxID=9606;

RN (1) NUCLEOTIDE SEQUENCE.

RP MEDLINE=99011253; PubMed=9792860;

RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,

RA Cochot P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;

RT "Determination of the genomic structure of the COL4A4 gene and of

RT novel mutations causing autosomal recessive Alport syndrome.";

RL Am. J. Hum. Genet. 63:1329-1340(1998).

RL EMBL; Y17397; CAA76763.1; -, Genomic_DNA.

DR EMBL; Y17441; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17440; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17442; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17398; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17399; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17443; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17400; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17402; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17411; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17410; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17409; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17408; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17407; CAA76763.1; JOINED; Genomic_DNA.

DR EMBL; Y17406; CAA76763.1; JOINED; Genomic_DNA.


```

DR EMBL; Y17405; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17404; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17403; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17428; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17427; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17426; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17425; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17424; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17423; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17422; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17421; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17420; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17439; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17438; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17437; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17436; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17435; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17434; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17433; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17432; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17431; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17430; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17429; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17419; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17418; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17417; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17416; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17415; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17414; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17413; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17412; CAA76763.1; JOINED; Genomic DNA.
DR EMBL; Y17401; CAA76763.1; JOINED; Genomic DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1q helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 20.
DR ProDom; PD000007; C1q helix; 9.
DR ProDom; PD003923; Procollagen4_C; 2.
DR SMART; SM00111; C4; 2.
DR Collagen.
KW SMART.
SQ SEQUENCE 1690 AA; 164096 MW; E1E72F283A72BAAE CRC64;

Query Match 88.1%; Score 37; DB 2; Length 1690;
Best Local Similarity 87.5%; Pred. No. 9.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    ||||| |||
Db 533 GAEGPPGL 540

RESULT 44
Q4RWM6 TETNG
ID Q4RWM6 TETNG PRELIMINARY; PRT; 1726 AA.
AC Q4RWM6
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Chromosome 3 SCAP14987, whole genome shotgun sequence.
GN ORFNames=GSTENG00027784001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.

Query Match 88.1%; Score 37; DB 2; Length 1690;
Best Local Similarity 87.5%; Pred. No. 9.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    ||||| |||
Db 533 GAEGPPGL 540

RESULT 44
Q4RWM6 TETNG
ID Q4RWM6 TETNG PRELIMINARY; PRT; 1726 AA.
AC Q4RWM6
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE Chromosome 3 SCAP14987, whole genome shotgun sequence.
GN ORFNames=GSTENG00027784001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

```

RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castellani L., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crollius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAE01014987; CAG07206.1; -; Genomic DNA.
SQ SEQUENCE 1726 AA; 187624 MW; E769ACC3815015EC CRC64;

Query Match 88.1%; Score 37; DB 2; Length 1726;
Best Local Similarity 87.5%; Pred. No. 9.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    ||||| |||
Db 1616 GAEGCPGL 1623

RESULT 45
Q5NC68 MOUSE
ID Q5NC68 MOUSE PRELIMINARY; PRT; 95 AA.
AC Q5NC68
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DE Procollagen, type XXIII, alpha 1 (Fragment).
GN Name=Col23a1; ORFNames=RP23-411E22.1-002;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Van Hellmond Z.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662843; CAI35930.1; -; Genomic DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1q helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 1.
DR ProDom; PD000007; C1q helix; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 95 AA; 9670 MW; 5CA553BE17FD3BAD CRC64;

Query Match 85.7%; Score 36; DB 2; Length 95;
Best Local Similarity 87.5%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    ||||| |||
Db 66 GAEGSPGL 73

RESULT 46

```

Q9CZS2 MOUSE
ID Q9CZS2_MOUSE PRELIMINARY; PRT; 150 AA.
AC Q9CZS2_MOUSE
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 11 days embryo whole body cDNA, RIKEN full-length
DE enriched library, clone:2700007f12 product:COLLAGEN ALPHA 1(XVI) CHAIN
DE homolog
GN Name=Coll16a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RX Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kiteunai T., Iashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,

RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771 (2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Whole body;
RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ono M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai K., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RC EMBL; AK012212; BAB28100.1; -; mRNA.
DR MGI; MGI:1095396; Coll16a1.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005515; P:protein binding; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen_2.
DR ProDom; PD000007; Clg_helix; 1.
KW Collagen.
SQ SEQUENCE 150 AA; 14823 MW; D956EF9160987FC8 CRC64;
Query Match 85.7%; Score 36; DB 2; Length 150;
Best Local Similarity 87.5%; Pred. No. 1.2e-02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
DB 34 GAEGSPGL 41
RESULT 47
Q61246_MOUSE PRELIMINARY; PRT; 187 AA.
ID Q61246_MOUSE
AC Q61246_MOUSE
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha 1 type XI collagen (Fragment).
GN Name=Coll1a1; Synonyms=moucaoll1a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BALB/c; TISSUE=Cartilage;
RX MEDLINE=97293217; PubMed=9164858;
RA Perselae M., Savontaus M., Metsaeranta M., Vuorio E.;
RT "Developmental regulation of mRNA species for types II, IX and XI
RT collagens during mouse embryogenesis.";
RL Biochem. J. 324:209-216 (1997).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BALB/c; TISSUE=Cartilage;
RA Perselae M.P.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
RC EMBL; X91014; CAA622496.1; -; mRNA.
DR MGI; MGI:88446; Coll1a1.
DR GO; GO:0001502; P:collagen condensation; IMP.
DR GO; GO:0030199; P:collagen fibril organization; IMP.

```

DR GO: 0006029; P: proteoglycan metabolism; IMP.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 4.
DR ProDom; PD000007; C1g_helix; 1.
KW Collagen.
FT NON_TER 1 187
FT NON_TER 187 187
SQ SEQUENCE 187 AA; 17261 MW; F4A43AA67CBBFF320 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 187;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 126 GAQGAPGL 133

RESULT 48
Q4NK30_9M1CC PRELIMINARY; PRT; 215 AA.
ID Q4NK30;
AC Q4NK30;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Putative membrane protein.
GN ORFNames=ArthDRAFT_3564;
OS Arthrobacter sp. FB24.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=290399;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FB24;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Copeland A., Lucas S., Lapidus A., Barry K., Detter C., Glavina T.,
RA Hamon N., Israni S., Pittluck S., Richardson P.;
RT "Sequencing of the draft genome assembly of Arthrobacter sp. FB24."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FB24;
RG US DOE Joint Genome Institute (JGI-PGF);
RA Larimer F., Land M.;
RT "Annotation of the draft genome assembly of Arthrobacter sp. FB24."
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAHQ01000002; EAL97697.1; -; Genomic DNA.
SQ SEQUENCE 215 AA; 21997 MW; FF9C6387A2A5F972 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 215;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 80 GAQGAPGL 87

RESULT 49
Q71RG9_HUMAN PRELIMINARY; PRT; 218 AA.
ID Q71RG9;
AC Q71RG9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE FP1572.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;

```

```

OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF370368; AA015204.1; -; mRNA.
DR Ensembl; ENSG0000084636; Homo sapiens.
DR GO: 0005737; P: cytoplasm; IEA.
DR GO: 0006817; P: phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 3.
DR ProDom; PD000007; C1g_helix; 1.
KW Collagen.
SQ SEQUENCE 218 AA; 21734 MW; 3A3C5B9194196715 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 218;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 102 GAEGSPGL 109

RESULT 50
Q6A622_PROAC PRELIMINARY; PRT; 287 AA.
ID Q6A622;
AC Q6A622;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Methylentetrahydrofolate dehydrogenase.
GN OrderedLocusNames=PPA1743;
OS Propionibacterium acnes.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacteriaceae; Propionibacteriaceae; Propionibacterium.
OX NCBI_TaxID=1747;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=KPA171202 / DSM 16379;
RX PubMed=15286373; DOI=10.1126/science.1100330;
RA Bruggemann H., Henne A., Hoster P., Liesegang H., Wierze A.,
RA Strittmatter A., Hujer S., Duerre P., Gottschalk G.;
RT "The complete genome sequence of Propionibacterium acnes, a commensal
RT of human skin."
RL Science 305:671-673 (2004).
DR EMBL; AE017283; AAT83472.1; -; Genomic DNA.
DR GO: 0003824; P: catalytic activity; IEA.
DR GO: 0009336; P: folic acid and derivative biosynthesis; IEA.
DR InterPro; IPR000672; THFDH/CYC_hydro.
DR Pfam; PF00763; THF DHG CYH; 1.
DR Pfam; PF02882; THF DHG CYH; 1.
DR PRINTS; PR00085; THFDH/DHGRASE.
DR ProDom; PD002300; THFDH/CYC_hydro; 1.
KW Complete proteome.
SQ SEQUENCE 287 AA; 30296 MW; B514D759213E2C8A CRC64;

Query Match 85.7%; Score 36; DB 2; Length 287;
Best Local Similarity 87.5%; Pred. No. 2.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 209 GAEGSPGL 216

RESULT 51
Q61TB7_CABER PRELIMINARY; PRT; 321 AA.
ID Q61TB7;
AC Q61TB7;

```

DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein CBG05818.
 GN Name=CBG05818;
 OS Caenorhabditis briggsae.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6238;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RG The C.briggsae Sequencing Consortium;
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CAAC01000028; CAE61844.1; -; Genomic_DNA.
 DR GO; GO:0005737; Cytoplasm; IEA.
 DR GO; GO:0042302; F:structural constituent of cuticle; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002486; Col cuticle_N.
 DR Pfam; PF01391; Collagen; 2.
 DR Pfam; PF01484; Col cuticle_N; 1.
 DR Collagen; Hypothetical protein.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 321 AA; 31851 MW; 619254D39110C46D CRC64;

 Query Match 85.7%; Score 36; DB 2; Length 321;
 Best Local Similarity 75.0%; Pred. No. 2.6e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 GAEGSPGL 8
 |||||
 Db 214 GAEGQPGM 221

 RESULT 52
 Q82LA4 STRAW
 ID Q82LA4 STRAW PRELIMINARY; PRT; 340 AA.
 AC Q82LA4;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN OrderedLocusNames=SAV2107;
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 RN [2]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=NA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 DR ENBL; BR000030; BAC69818.1; -; Genomic DNA.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR01087; Lipase_GDSL.
 DR Pfam; PF00657; Lipase_GDSL; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 340 AA; 35786 MW; 38B38C5AE076F3A0 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 340;
 Best Local Similarity 75.0%; Pred. No. 2.7e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GAEGSPGL 8
 |||||
 Db 99 GADTFGL 106

 RESULT 53
 Q7G6D9 ORYSA
 ID Q7G6D9 ORYSA PRELIMINARY; PRT; 374 AA.
 AC Q7G6D9;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Hypothetical protein OSJNAA0034B05.10.
 GN ORFNames=OSJNAA0034B05.10;
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Wing R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
 RA Rambo T., Saski C., Henry D., Oates R., Simmons J.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice
 RT chromosome 10.";
 RL Science 300:1566-1569(2003).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC122145; AAM46055.1; -; Genomic DNA.
 DR EMBL; AE017058; AAP52194.1; -; Genomic_DNA.
 DR Gramene; Q7G6D9; -;
 DR GO; GO:0005529; F:sugar binding; IEA.
 DR InterPro; IPR001480; B_lectin.
 KW Hypothetical protein.
 SQ SEQUENCE 374 AA; 36305 MW; D262F66A475A8F5C CRC64;

 Query Match 85.7%; Score 36; DB 2; Length 374;
 Best Local Similarity 87.5%; Pred. No. 3e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

 QY 1 GAEGSPGL 8
 |||||
 Db 189 GAGSPGL 196

 RESULT 54
 Q8SB75 ORYSA
 ID Q8SB75 ORYSA PRELIMINARY; PRT; 374 AA.
 AC Q8SB75;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein OSJNBA0004E08.6.
 GN Name=OSJNBA0004E08.6;
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
 RA Saski C., Henry D., Oates R., Simmons J., Wilson R., Minx P.,

```
RA Kyung K.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC091724; AAL75733.1; -; Genomic_DNA.
DR Gramene; Q8SB75; -.
DR GO; GO:0005529; F:sugar binding; IEA.
DR InterPro; IPR001480; B_lectin.
KW Hypothetical protein.
SQ SEQUENCE 374 AA; 36305 MW; D262F66A475A8F5C CRC64;

Query Match 85.7%; Score 36; DB 2; Length 374;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 189 GAGGSPGL 196

RESULT 55
Q568Y4 RAT PRELIMINARY; PRT; 415 AA.
AC Q568Y4;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Coll6al, predicted protein.
GN Name=Coll6al, predicted;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Rouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC092654; AAH92654.1; -; mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
KW Collagen.
SQ SEQUENCE 415 AA; 39879 MW; 693F7D948237D1AA CRC64;

Query Match 85.7%; Score 36; DB 2; Length 415;
Best Local Similarity 87.5%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 189 GAGGSPGL 196
```

```
Db 299 GAGGSPGL 306

RESULT 56
Q16593 HUMAN PRELIMINARY; PRT; 447 AA.
AC Q16593;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 25, Last annotation update)
DE Collagen-like protein (447 AA) (fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP TISSUE=Placenta;
RA Kimura S.;
RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.
DR EMBL; X14963; CAA33085.1; -; mRNA.
DR EMBL; X15038; CAA33142.1; -; mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; C1g_helix; 1.
KW Collagen.
FT NON_TER 1 1
FT NON_TER 447 447
SQ SEQUENCE 447 AA; 41829 MW; FDB207023D87CC94 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 447;
Best Local Similarity 75.0%; Pred. No. 3.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||| |||||
Db 419 GAGGSPGV 426

RESULT 57
EGR4 HUMAN STANDARD; PRT; 486 AA.
AC Q05215;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Early growth response protein 4 (EGR-4) (AT133).
GN Name=EGR4;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
[1]
RN NUCLEOTIDE SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=93278383; PubMed=8504297;
RA Holst C., Skerka C., Lichter P., Bialonski A., Zipfel P.F.;
RT "Genomic organization, chromosomal localization and promoter function
RT of the human zinc-finger gene pAT133.";
RL Hum. Mol. Genet. 2:367-372(1993).
[2]
RN NUCLEOTIDE SEQUENCE.
RX MEDLINE=92052214; PubMed=1658795;
RA Mueller H.-J., Skerka C., Bialonski A., Zipfel P.F.;
RT "Clone pAT133 identifies a gene that encodes another human member of a
RT class of growth factor-induced genes with almost identical zinc-finger
RT domains.";
```

Proc. Natl. Acad. Sci. U.S.A. 88:10079-10083(1991).
CC - FUNCTION: Transcriptional regulator (By similarity).
CC - SUBCELLULAR LOCATION: Nuclear (Probable).
CC - INDUCTION: By PHA/PMA or by serum.
CC - SIMILARITY: Belongs to the EGR C2H2-type zinc-finger protein family.
CC - SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
CC -----
CC EMBL: X69438; CA94214.1; -; Genomic_DNA.
CC EMBL: X60104; CA942698.1; -; mRNA.
CC PIR: A41537; A41537.
CC HSSP: P08046; 1P47.
CC SMR: Q05215; 378-463.
CC TRANSFAC: T05190; -.
CC Ensembl: ENSG00000135625; Homo sapiens.
CC HGNC: HGNC:3241; EGR4.
CC MIM: 128992; -.
CC GO: GO:0003700; F:transcription factor activity; NAS.
CC GO: GO:0008284; P:positive regulation of cell proliferation; TAS.
CC InterPro: IPR007087; Znf_C2H2.
CC Pfam: PF00096; zf-C2H2; 3.
CC ProDom: PD000003; Znf_C2H2; 2.
CC SMART: SM00355; Znf_C2H2; 3.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; 3.
CC DNA-binding; Metal-binding; Nuclear protein; Repeat; Transcription;
KW Transcription regulation; Zinc; Zinc-finger.
KW ZN_FING 380 404 C2H2-type 1.
FT ZN_FING 410 432 C2H2-type 2.
FT ZN_FING 438 460 C2H2-type 3.
FT CONFLICT 427 427 S -> T (in Ref. 2).
SQ SEQUENCE 486 AA; 50856 MW; 0DF764427E0A21E3 CRC64;

Query Match 85.7%; Score 36; DB 1; Length 486;
Best Local Similarity 87.5%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 279 GAEGLPGL 286

RESULT 58
Q5TKQ7 ORYSA
ID Q5TKQ7 ORYSA PRELIMINARY; PRT; 531 AA.
AC Q5TKQ7;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein OJ1362_G11.1 (Hypothetical protein OSJNB0052K01.22).
GN Names:OJ1362_G11.1; Synonyms:cultivar-group).
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriaraloideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M., Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R., Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H., Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L., Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W., Wu H.-P., Shaw J.-F.;
RT "Oryza sativa BAC OJ1362_G11 genomic sequence."
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.

[2]
RN NUCLEOTIDE SEQUENCE.
RP Chow T.-Y., Hsing Y.-I.C., Chen C.-S., Chen H.-H., Liu S.-M., Chao Y.-T., Chang S.-J., Chen H.-C., Chen S.-K., Chen T.-R., Chen Y.-L., Cheng C.-H., Chung C.-I., Han S.-Y., Hsiao S.-H., Hsiung J.-N., Hsu C.-H., Huang J.-J., Kau P.-I., Lee M.-C., Leu H.-L., Li Y.-F., Lin S.-J., Lin Y.-C., Wu S.-W., Yu C.-Y., Yu S.-W., Wu H.-P., Shaw J.-F.;
RT "Oryza sativa BAC OSJNB0052K01 genomic sequence."
RL Submitted (NOV-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC104713; AAV67810.1; -; Genomic_DNA.
DR EMBL: AC119291; AAV59412.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 531 AA; 60523 MW; 0B3EA2853A41DDAB CRC64;

Query Match 85.7%; Score 36; DB 2; Length 531;
Best Local Similarity 87.5%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 166 GAEGSPGL 173

RESULT 59
Q8K036 MOUSE
ID Q8K036 MOUSE PRELIMINARY; PRT; 565 AA.
AC Q8K036;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Coll3a1 protein.
GN Name:Coll3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC034164; AAH34164.1; -; mRNA.
DR MGI: MGI:1277201; Coll3a1.
DR GO: GO:0005911; C:intercellular junction; IDA.
DR InterPro: IPR008161; Clg_helix.

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-FVB/N; TISSUE=Mammary tumor. C3;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023940; AAH23940.1; -; mRNA.
DR Ensembl; ENSMUSG0000040690; Mus musculus.
DR MGI; MGI:1095396; Coll16a1.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 575 AA; 54847 MW; 0B5420D04086A313 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 575;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 459 GAGSPGL 466

RESULT 62
Q86TR4 HUMAN PRELIMINARY; PRT; 598 AA.
AC Q86TR4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Full-length cDNA 5-PRIME end of clone CS0DI001YE04 of Placenta of Homo sapiens (human) (Fragment).
DE sapiens (human) (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RA Li W.B., Gruber C., Jesse J., Polayes D.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Placenta;
RA Genoscope;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX248768; CAD66575.1; -; mRNA.
DR Ensembl; ENSG00000165801; Homo sapiens.

FT NON_TER 1 1
FT NON_TER 598 598
SQ SEQUENCE 598 AA; 63461 MW; 8CA13483D4831BF6 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 598;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AEGSPGL 8
DB 300 AEGSPGL 306

RESULT 63
O68220 ANAPH PRELIMINARY; PRT; 619 AA.
AC O68220;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE 130kDa protein.
GN Name-rea;
OS Anaplasma phagocytophilum (Ehrlichia phagocytophila).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Anaplasmataceae; Anaplasma; phagocytophilum group.
OX NCBI_TaxID=948;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=USG3;
RX MEDLINE=98187902; PubMed=9529053;
RA Storey J.R., Doros-Richert L.A., Gingrich-Baker C., Munroe K., Mather T.N., Coughlin R.T., Beltz G.A., Murphy C.I.;
RT "Molecular cloning and sequencing of three granulocytic Ehrlichia genes encoding high-molecular-weight immunoreactive proteins.";
RL Infect. Immun. 66:1356-1363 (1998).
DR EMBL; AF020522; AAC12863.1; -; Genomic DNA.
SQ SEQUENCE 619 AA; 66109 MW; 60789E4A8DA11553 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 5.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AEGSPGL 8
DB 15 AEGSPGL 21

RESULT 64
O17866 CAEBL PRELIMINARY; PRT; 650 AA.
AC O17866;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE CLE-1C protein (Hypothetical protein cle-1).
GN Name=cle-1; ORFNames=C36B1.1, C36B1.1c;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae; Rhabditidae.
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21157401; PubMed=11257122; DOI=10.1083/jcb.152.6.1219;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J., Kramer J.M.;
RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance.";
RL J. Cell Biol. 152:1219-1232 (2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology.",

RL Science 282:2012-2018(1998).

DR EMBL; AF164959; AAD47825.1; -; Genomic DNA.

DR EMBL; Z81079; CAB03084.1; -; Genomic DNA.

DR PIR; T22002; T22002.

DR HSSP; P39061; 1KOE.

DR Ensembl; C36B1.1; Caenorhabditis elegans.

DR WormBase; WBGene0000527; cle-1.

DR WormPep; C36B1.1c; CEI7816.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0031012; C:extracellular matrix; IEA.

DR GO; GO:0005198; P:structural molecule activity; IEA.

DR GO; GO:0007155; P:cell adhesion; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR010515; Endostatin.

DR Pfam; PF01391; Collagen; 2.

DR Pfam; PF06482; Endostatin; 1.

KW Collagen; Complete procollagen; Hypothetical protein.

SQ SEQUENCE 650 AA; 69597 MW; 6CF29ED9C16B170E CRC64;

Query Match 85.7%; Score 36; DB 2; Length 650;

Best Local Similarity 75.0%; Pred. No. 5.3e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 209 GADGAPGL 216

|||||

RESULT 65

O93486 ONCMY

ID O93486 ONCMY PRELIMINARY; PRT; 678 AA.

AC O93486;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Alpha 3 type I collagen (Fragment).

OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;

OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.

OK NCBI_TaxID=8022;

FN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Fibroblast;

RA Saito M., Kunisaki N., Hirono I., Aoki T., Ishida M., Urano N.,

RA Kimura S.,

RT "Partial characterization of cDNA clones encoding the three distinct

RT pro alpha chains of type I collagen from rainbow trout.",

RL Fish. Sci. 64:780-786(1998).

DR EMBL; AB008374; BAA33381.1; -; mRNA.

DR GO; GO:0005581; C:collagen; IEA.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0005201; P:extracellular matrix structural constituent; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008160; Collagen.

DR InterPro; IPR000885; Fib_collagen_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 6.

DR ProDom; PD002078; Fib_collagen_C; 1.

DR SMART; SM00038; COLFI; 1.

KW Collagen.

FT NON TER

SQ SEQUENCE 678 AA; 66130 MW; 87CD72BE2B41C1E9 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 678;

Best Local Similarity 87.5%; Pred. No. 5.6e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

|||||

Db 389 GAGSPGL 396

RESULT 66

O70575 MOUSE

ID O70575 MOUSE PRELIMINARY; PRT; 739 AA.

AC O70575;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Collagen type XIII alpha-1 chain.

GN Name=Coll3a1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OK NCBI_TaxID=10090;

FN [1]

RP NUCLEOTIDE SEQUENCE.

RC TISSUE=Gut;

RX MEDLINE=98288296; PubMed=9624150; DOI=10.1074/jbc.273.25.15590;

RA Hagg P., Rehn M., Huhtala P., Vaisanen T., Tamminen M.,

RA Pihlajaniemi T.;

RT "Type XIII collagen is identified as a plasma membrane protein.";

RL J Biol. Chem. 273:15590-15597(1998).

DR EMBL; U30292; AAC24314.1; -; mRNA.

DR MGI; MGI:1277201; Coll3a1.

DR GO; GO:0005911; C:intercellular junction; IDA.

DR InterPro; IPR008161; Clg_helix.

DR InterPro; IPR008160; Collagen.

DR Pfam; PF01391; Collagen; 9.

DR ProDom; PD000007; Clg_helix; 3.

KW Collagen.

SQ SEQUENCE 739 AA; 72110 MW; F13951061381F017 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 739;

Best Local Similarity 87.5%; Pred. No. 6.1e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 363 GAGSPGL 370

|||||

RESULT 67

O7D974 MYCTU

ID O7D974 MYCTU PRELIMINARY; PRT; 749 AA.

AC O7D974;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

DE PE PGSR family protein.

GN OrderedLocusNames=MT0854.1;

OS Mycobacterium tuberculosis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;

OK NCBI_TaxID=1773;

FN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=CDC 1551 / Oshkosh;

RX MEDLINE=22206494; PubMed=12218036;

RX DOI=10.1128/JB.184.19.5479-5490.2002;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,

RA Peterson J.D., DeBoy R.T., Dodson R.J., Gwinn M.L., Haft D.H.,

RA Hickey E.K., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D.,

RA Salzberg S.L., Delcher A., Utterback T.R., Weidman J.F., Khouri H.M.,

RA Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C.,

RA Fraser C.M.;

RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and

RT laboratory strains.",

RL J. Bacteriol. 184:5479-5490(2002).

DR EMBL; AE000516; AAK45096.1; -; Genomic DNA.

```
DR TIGR: MT0854.1; -;
SQ SEQUENCE 749 AA; 57432 MW; 9B21A1CD252936FA CRC64;

Query Match 85.7%; Score 36; DB 2; Length 749;
Best Local Similarity 87.5%; Pred. NO. 6.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 740 GAEGLPGL 747

RESULT 68
Q9RFV7 MYCTU PRELIMINARY; PRT; 749 AA.
AC Q9RFV7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE PE-PGRS FAMILY PROTEIN.
GN Name=PE_PGRS13; OrderedLocusNames=Rv0833;
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1773;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=96342230; DOI=10.1038/31159;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C.M.,
RA Harris D.E., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,
RA Teklaia F., Badcock K., Basham D., Brown D., Chillingworth T.,
RA Connor R., Davies R.M., Devlin K., Feltwell T., Gentles S., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S.,
RA Murphy J.D., Oliver S., Osborne J., Quail M.A., Rajandream M.A.,
RA Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; BX842574; CAE55328.1; -; Genomic_DNA.
DR TubercuList; Rv0833; -.
KW Complete proteome.
SQ SEQUENCE 749 AA; 57448 MW; C3C91307F9A8A2EC CRC64;

Query Match 85.7%; Score 36; DB 2; Length 749;
Best Local Similarity 87.5%; Pred. NO. 6.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 740 GAEGLPGL 747

RESULT 69
Q9RIN9 MOUSE PRELIMINARY; PRT; 751 AA.
AC Q9RIN9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type XIII collagen.
GN Name=Coll3a1; Synonym=coll3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=99357014; PubMed=10429945; DOI=10.1016/S0945-053X(99)00018-9;
RX Kvist A.P., Latvanlehto A., Sund M., Horelli-Kuitunen N., Rehn M.,
RA
```

```
RA Palotie A., Beier D., Pihlajaniemi T.;
RT "Complete exon-intron organization and chromosomal location of the
RT gene for mouse type XIII collagen (coll3a1) and comparison with its
RT human homologue.";
RL Matrix Biol. 18:261-274(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Kvist A.-P., Latvanlehto A., Horelli-Kuitunen N., Sund M., Rehn M.,
RA Beier D.R., Palotie A., Pihlajaniemi T.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF063693; AAD50327.1; -; Genomic_DNA.
DR EMBL; AF063666; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063668; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063667; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063669; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063671; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063673; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063675; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063677; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063686; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063685; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063684; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063683; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063682; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063681; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063680; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063679; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063678; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063692; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063691; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063690; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063689; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063688; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063687; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063676; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063674; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063672; AAD50327.1; JOINED; Genomic_DNA.
DR EMBL; AF063670; AAD50327.1; JOINED; Genomic_DNA.
DR Ensembl; ENSMUSG0000058806; Mus musculus.
MGI; MGI:1277201; Coll3a1.
GO; GO:0005911; C:intercellular junction; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 9.
DR ProDom; PD0000007; Clg_helix; 3.
KW Collagen.
SQ SEQUENCE 751 AA; 73172 MW; FB2443E1CBF51AD CRC64;

Query Match 85.7%; Score 36; DB 2; Length 751;
Best Local Similarity 87.5%; Pred. NO. 6.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
DB 363 GAEGSPGL 370

RESULT 70
Q7U160 MYCBO PRELIMINARY; PRT; 773 AA.
AC Q7U160;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PE-PGRS FAMILY PROTEIN.
GN Name=PE_PGRS13; OrderedLocusNames=Mb0856;
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex.
OX NCBI_TaxID=1765;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```

RC STRAIN=AF2122/97; PubMed=12788972; DOI=10.1073/pnas.1130426100;
RX MEDLINE=22703107;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
DR EMBL; BX248336; CAD93718.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 773 AA; 59296 MW; 4ECCDAD879A7B822 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 773;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 764 GAEGLPGL 771

RESULT 71
Q9U9K6 CAEL PRELIMINARY; PRT; 778 AA.
AC Q9U9K6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE CLF-1B protein (Hypothetical protein cle-1).
GN Name=cle-1; ORFName=C36B1.1, C36B1.1B;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21157401; PubMed=11257122; DOI=10.1083/jcb.152.6.1219;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,
RA Kramer J.M.;
RT "The NCL1/endostatin domain of Caenorhabditis elegans type XVIII
RT collagen affects cell migration and axon guidance.";
RL J. Cell Biol. 152:1219-1232 (2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; AF164959; AAD47824.1; -; Genomic DNA.
DR EMBL; Z80215; CAD21700.1; -; Genomic DNA.
DR EMBL; Z81079; CAD21643.1; -; Genomic DNA.
DR EMBL; Z81079; CAD21700.1; JOINED; Genomic DNA.
DR EMBL; Z80215; CAD21643.1; JOINED; Genomic DNA.
DR HSSP; P39061; IKOE.
DR Ensemble; C36B1.1; Caenorhabditis elegans.
DR WormBase; WBGene0000527; cle-1.
DR WormPep; C36B1.1b; CE29706.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0005578; Extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF06482; Endostatin; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Complete proteome; Extracellular matrix;
KW Hypothetical protein; Structural protein.
SQ SEQUENCE 778 AA; 83788 MW; 743A7FC8A7FF1ABD CRC64;

```

```

Query Match 85.7%; Score 36; DB 2; Length 778;
Best Local Similarity 75.0%; Pred. No. 6.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 337 GADGAPGL 344

RESULT 72
COB1 BOVIN STANDARD; PRT; 911 AA.
ID COB1 BOVIN STANDARD; PRT; 911 AA.
AC Q28083;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XI) chain precursor (Fragment).
GN Name=COL1A1;
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
OC Pecora; Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX TISSUE=Smooth muscle;
RX MEDLINE=92078200; PubMed=1741123;
RA Brown K.E., Lawrence R., Sonenshein G.E.;
RT "Concerted modulation of alpha 1(XI) and alpha 2(V) collagen mRNAs in
RT bovine vascular smooth muscle cells.";
RL J. Biol. Chem. 266:23268-23273 (1991).
RN [2]
RP PROTEIN SEQUENCE OF 347-354; 356-363 AND 586-600.
RC TISSUE=Eye vitreous humor;
RX PubMed=8486632;
RA Mayne R., Brewton R.G., Mayne P.M., Baker J.R.;
RT "Isolation and characterization of the chains of type V/type XI
RT collagen present in bovine vitreous.";
RL J. Biol. Chem. 268:9381-9386 (1993).
CC -!- FUNCTION: May play an important role in fibrillogenesis by
CC controlling lateral growth of collagen II fibrils.
CC -!- SUBUNIT: Trimers composed of three different chains: alpha 1(XI),
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
CC modification of alpha 1(XI). Alpha 1(XI) can also be found instead
CC of alpha 3(XI)=1(XI) (By similarity).
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
-----
DR EMBL; M82977; AAA30369.1; -; mRNA.
DR PIR; S18251; S18251.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 1.
DR ProDom; PD000007; C1g_helix; 4.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Hydroxylation; Repeat; Structural protein.
FT PROPEP <1 278 N-terminal propeptide (Potential).
FT CHAIN 279 >911 Collagen alpha 1(XI) chain.
FT REGION <1 186 Nonhelical region.
FT REGION 187 275 Triple-helical region (interrupted).
FT REGION 276 278 Short nonhelical segment.
FT REGION 279 295 Telopeptide.
FT REGION 296 >911 Triple-helical region.
FT MOD_RES 379 379 Allylsine.
FT NON_TER 1

```

```

FT  NON TER      911      911
SQ  SEQUENCE     911 AA;  89260 MW;  C05C4B3350749CFC CRC64;

Query Match      85.7%; Score 36; DB 1; Length 911;
Best Local Similarity 75.0%; Pred. No. 7.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
DB      803 GAGGAPGL 810

RESULT 73
Q6ZQK3_MOUSE PRELIMINARY; PRT; 915 AA.
AC  Q6ZQK3
DT  05-JUL-2004 (TrEMBLrel. 27, Created)
DT  05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT  05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE  MKIAA0049 protein (Fragment).
GN  Name=Nbr1; Synonyms=MKIAA0049;
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC  Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RG  TISSUE=Embryonic tail;
RX  PubMed=14621295;
RA  Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA  Saga Y., Nagase T., Ohara O., Koga H.;
RT  Prediction of the coding sequences of mouse homologues of KIAA gene:
RT  i11. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT  cDNAs identified by screening of terminal sequences of cDNA clones
RT  randomly sampled from size-fractionated libraries.";
RL  DNA Res. 10:167-180(2003).
DR  EMBL; AK129043; BAC97853.1; -; mRNA.
DR  SMR; Q6ZQK3; 7-95.
DR  MGI; MGI:108499; Nbr1.
DR  GO; GO:0008270; F:zinc ion binding; IEA.
DR  InterPro; IPR000270; OPR_PBI.
DR  InterPro; IPR000433; Znf_ZZ.
DR  Pfam; PF00564; PBI; 1.
DR  SMART; SM00666; PBI; 1.
DR  SMART; SM00291; Znf_ZZ; 1.
DR  PROSITE; PS01357; ZF_ZZ_1; UNKNOWN_1.
DR  PROSITE; PS0135; ZF_ZZ_2; 1.
FT  NON TER      1
SQ  SEQUENCE     915 AA; 101813 MW;  86C520F6463D2D82 CRC64;

Query Match      85.7%; Score 36; DB 2; Length 915;
Best Local Similarity 75.0%; Pred. No. 7.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
DB      790 GAEGEFGI 797

RESULT 74
Q4SIU4_TETNG PRELIMINARY; PRT; 925 AA.
AC  Q4SIU4;
DT  13-SEP-2005 (TrEMBLrel. 31, Created)
DT  13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT  13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE  Chromosome 21 SCAR14577, whole genome shotgun sequence.
DE  (fragment).
GN  ORFNAMES=GSTENG00017515001;
OS  Tetraodon nigroviridis (Green puffer).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

```

OC  Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC  Tetraodontidae; Tetraodontidae; Tetraodon.
OX  NCBI_TaxID=99883;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RA  Jallion O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA  Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA  Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA  Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA  Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA  Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA  Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA  Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
RA  Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,
RA  Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA  Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA  Winkler P., Lander E.S., Weissenbach J., Roest Croallius H.;
RT  "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT  the early vertebrate proto-karyotype.";
RL  Nature 431:946-957(2004).
RN  [2]
RP  NUCLEOTIDE SEQUENCE.
RG  Genoscope; Whitehead Institute Centre for Genome Research;
RL  Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC  -!- CAUTION: The sequence shown here is derived from an
CC  EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC  preliminary data.
DR  EMBL; CAAE01014577; CAF99438.1; -; Genomic_DNA.
DR  InterPro; IPR008161; C1g_helix.
DR  InterPro; IPR008160; Collagen.
DR  Pfam; PF01391; Collagen; 5.
DR  ProDom; PD000007; C1g_helix; 4.
KW  Collagen.
FT  NON TER      1
SQ  SEQUENCE     925 AA; 93961 MW;  4E184A0B1723BF26 CRC64;

Query Match      85.7%; Score 36; DB 2; Length 925;
Best Local Similarity 87.5%; Pred. No. 7.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
DB      796 GAGGSPGL 803

RESULT 75
Q501R9_RAT PRELIMINARY; PRT; 983 AA.
AC  Q501R9;
DT  13-SEP-2005 (TrEMBLrel. 31, Created)
DT  13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT  13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE  Hypothetical protein RGD1311421_predicted.
GN  Name=RGD1311421_predicted;
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC  Muridae; Murinae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  NUCLEOTIDE SEQUENCE.
RC  TISSUE=Lung;
RX  MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA  Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max J.I., Wang J., Hsieh F.,
RA  Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Udesin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

```

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Lung;
 RG NIH MGC Project;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC095903; AAH95903.1; -; mRNA.
 DR SMR; Q501R9; 1-86.
 DR InterPro; IPR000270; OPR_PBI.
 DR InterPro; IPR000449; UBA_
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00564; PBI; 1.
 DR SMART; SM00666; PBI; 1.
 DR SMART; SM00291; Znf_ZZ; 1.
 DR PROSITE; PS00303; UBA; 1.
 DR PROSITE; PS01357; ZF_ZZ_1; 1.
 DR PROSITE; PS0135; ZF_ZZ_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 983 AA; 109829 MW; 4A80880D9EDDE729 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 983;
 Best Local Similarity 75.0%; Pred. No. 8.2e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
 Db 781 GAEGEPGI 788

Search completed: March 11, 2006, 12:05:41
 Job time : 143.2 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:06:07 ; Search time 26.4 Seconds
(without alignments)
25.053 Million cell updates/sec

Title: US-10-698-121A-1
Perfect score: 42
Sequence: 1 GAEGSPGL 8

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5 COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/6 COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/H COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/PCTRUS COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/RE COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	92.9	495	2	US-09-252-991A-31949 Sequence 31949, A
2	38	90.5	219	2	US-10-153-469A-44 Sequence 44, Appl
3	38	90.5	219	2	US-10-153-469A-46 Sequence 46, Appl
4	38	90.5	219	2	US-10-104-889-44 Sequence 44, Appl
5	38	90.5	219	2	US-10-104-889-46 Sequence 46, Appl
6	38	90.5	238	2	US-09-252-991A-16703 Sequence 16703, A
7	38	90.5	926	2	US-09-252-991A-32551 Sequence 32551, A
8	38	90.5	1057	2	US-08-931-820-1 Sequence 1, Appli
9	38	90.5	1057	2	US-10-153-469A-20 Sequence 20, Appl
10	38	90.5	1057	2	US-10-104-889-20 Sequence 20, Appl
11	38	90.5	1107	2	US-10-153-469A-11 Sequence 11, Appl
12	38	90.5	1107	2	US-10-104-889-11 Sequence 11, Appl
13	38	90.5	1171	2	US-10-153-469A-8 Sequence 8, Appli
14	38	90.5	1171	2	US-10-104-889-8 Sequence 8, Appli
15	38	90.5	1341	2	US-08-963-825-18 Sequence 18, Appl
16	38	90.5	1341	2	US-09-500-811-18 Sequence 18, Appl
17	38	90.5	1341	2	US-09-570-573-18 Sequence 18, Appl
18	38	90.5	1341	2	US-09-548-608-18 Sequence 18, Appl
19	38	90.5	1388	2	US-10-153-469A-10 Sequence 10, Appl
20	38	90.5	1388	2	US-10-104-889-10 Sequence 10, Appl
21	38	90.5	1461	2	US-09-585-887-9 Sequence 9, Appli
22	38	90.5	1461	2	US-09-289-578-9 Sequence 9, Appli
23	38	90.5	1464	2	US-09-331-347C-21 Sequence 21, Appl
24	37	88.1	1027	2	US-10-360-101-221 Sequence 221, App
25	37	88.1	1218	2	US-09-949-016-7065 Sequence 7065, App
26	37	88.1	1590	2	US-09-949-016-5884 Sequence 5884, Ap
27	36	85.7	376	2	US-09-902-540-12854 Sequence 12854, A

28	36	85.7	484	2	US-09-066-046-8 Sequence 8, Appli
29	36	85.7	486	2	US-09-538-092-1269 Sequence 1269, Ap
30	36	85.7	486	2	US-09-949-016-6151 Sequence 6151, Ap
31	36	85.7	510	2	US-09-949-016-10733 Sequence 10733, A
32	36	85.7	619	2	US-09-066-046-2 Sequence 2, Appli
33	36	85.7	1603	2	US-09-949-016-6136 Sequence 6136, Ap
34	36	85.7	1609	2	US-09-949-016-10910 Sequence 10910, A
35	36	85.7	1745	2	US-09-795-061-4 Sequence 4, Appli
36	36	85.7	1745	2	US-09-949-002-405 Sequence 405, App
37	36	85.7	1771	2	US-09-949-002-492 Sequence 492, App
38	35	85.7	1806	2	US-09-919-497-56 Sequence 56, Appl
39	35	83.3	62	1	US-07-609-716-61 Sequence 61, Appl
40	35	83.3	62	1	US-08-642-255-31 Sequence 31, Appl
41	35	83.3	62	2	US-08-475-411A-61 Sequence 61, Appl
42	35	83.3	62	2	US-08-478-029A-61 Sequence 111, App
43	35	83.3	69	1	US-07-609-716-111 Sequence 111, App
44	35	83.3	69	2	US-08-475-411A-111 Sequence 111, App
45	35	83.3	69	2	US-08-478-029A-111 Sequence 111, App
46	35	83.3	72	1	US-07-609-716-113 Sequence 113, App
47	35	83.3	72	2	US-08-475-411A-113 Sequence 113, App
48	35	83.3	72	2	US-08-478-029A-113 Sequence 113, App
49	35	83.3	82	1	US-07-609-716-114 Sequence 114, App
50	35	83.3	85	2	US-08-475-411A-114 Sequence 114, App
51	35	83.3	85	2	US-08-478-029A-114 Sequence 114, App
52	35	83.3	115	2	US-09-902-540-10883 Sequence 10883, A
53	35	83.3	313	2	US-08-478-029A-113 Sequence 66, Appl
54	35	83.3	357	1	US-09-949-016-9265 Sequence 33, Appl
55	35	83.3	357	1	US-08-642-255-33 Sequence 66, Appl
56	35	83.3	357	2	US-08-475-411A-66 Sequence 66, Appl
57	35	83.3	357	2	US-08-478-029A-66 Sequence 66, Appl
58	35	83.3	377	2	US-09-252-991A-20479 Sequence 20479, A
59	35	83.3	430	2	US-09-902-540-13248 Sequence 13248, A
60	35	83.3	595	2	US-09-219-849-48 Sequence 48, Appl
61	35	83.3	595	2	US-09-219-849-50 Sequence 50, Appl
62	35	83.3	623	2	US-03-029-348-3 Sequence 3, Appli
63	35	83.3	626	2	US-09-029-348-2 Sequence 2, Appli
64	35	83.3	822	2	US-09-219-849-49 Sequence 49, Appl
65	35	83.3	825	2	US-10-210-428-1 Sequence 1, Appli
66	35	83.3	825	2	US-10-237-551-161 Sequence 161, App
67	35	83.3	826	2	US-08-994-998A-47 Sequence 47, Appl
68	35	83.3	826	2	US-10-237-551-47 Sequence 47, Appl
69	35	83.3	833	2	US-09-949-016-9530 Sequence 9530, Ap
70	35	83.3	965	2	US-09-949-016-6669 Sequence 6669, Ap
71	35	83.3	1057	2	US-08-931-820-4 Sequence 4, Appli
72	35	83.3	1057	2	US-10-153-469A-16 Sequence 16, Appl
73	35	83.3	1057	2	US-10-104-889-16 Sequence 16, Appl
74	35	83.3	1078	2	US-08-963-825-21 Sequence 21, Appl
75	35	83.3	1078	2	US-09-500-811-21 Sequence 21, Appl
76	35	83.3	1078	2	US-09-570-573-21 Sequence 21, Appl
77	35	83.3	1078	2	US-09-548-608-21 Sequence 21, Appl
78	35	83.3	1169	2	US-10-153-469A-6 Sequence 6, Appli
79	35	83.3	1169	2	US-10-104-889-6 Sequence 6, Appli
80	35	83.3	1670	2	US-09-949-016-5883 Sequence 5883, Ap
81	34	81.0	10	2	US-08-159-339A-898 Sequence 898, App
82	34	81.0	36	1	US-08-555-669-10 Sequence 10, Appl
83	34	81.0	36	2	US-09-073-663-10 Sequence 10, Appl
84	34	81.0	65	2	US-09-513-999C-7428 Sequence 7428, Ap
85	34	81.0	111	2	US-09-270-767-57930 Sequence 57930, A
86	34	81.0	163	2	US-10-000-489-58 Sequence 58, Appl
87	34	81.0	176	2	US-09-252-991A-20852 Sequence 20852, A
88	34	81.0	182	2	US-09-270-767-42615 Sequence 42615, A
89	34	81.0	293	1	US-08-907-492A-2 Sequence 2, Appli
90	34	81.0	295	1	US-08-907-492A-2 Sequence 2, Appli
91	34	81.0	355	2	US-09-786-240-13 Sequence 13, Appl
92	34	81.0	495	1	US-09-252-991A-27426 Sequence 27426, A
93	34	81.0	495	1	US-08-794-795-2 Sequence 2, Appli
94	34	81.0	495	2	US-09-249-200-2 Sequence 2, Appli
95	34	81.0	520	1	US-08-794-795-6 Sequence 6, Appli
96	34	81.0	520	2	US-09-249-200-6 Sequence 6, Appli
97	34	81.0	520	2	US-09-949-002-392 Sequence 392, App
98	34	81.0	520	2	US-09-949-002-553 Sequence 553, App
99	34	81.0	520	2	US-09-999-833A-614 Sequence 614, App
100	34	81.0	520	2	US-10-020-445A-614 Sequence 614, App

101	34	81.0	571	2	US-10-104-047-3814	Sequence 3814, Ap	174	32	76.2	164	1	US-08-154-915-6	Sequence 6, Appl
102	34	81.0	684	1	US-08-555-669-12	Sequence 12, Appl	175	32	76.2	164	1	US-08-375-983B-4	Sequence 4, Appl
103	34	81.0	684	2	US-09-073-663-12	Sequence 12, Appl	176	32	76.2	164	1	US-08-406-248-2	Sequence 2, Appl
104	34	81.0	689	2	US-09-949-016-11276	Sequence 11276, A	177	32	76.2	164	1	US-08-574-043A-2	Sequence 2, Appl
105	34	81.0	1017	2	US-08-468-996-10	Sequence 10, Appl	178	32	76.2	164	1	US-08-524-218A-2	Sequence 2, Appl
106	34	81.0	1060	2	US-08-931-820-3	Sequence 3, Appl	179	32	76.2	164	1	US-08-795-015-2	Sequence 2, Appl
107	34	81.0	1268	2	US-09-949-016-7487	Sequence 7487, Ap	180	32	76.2	164	2	US-08-456-297-2	Sequence 2, Appl
108	34	81.0	1418	2	US-08-963-825-20	Sequence 20, Appl	181	32	76.2	164	2	US-08-794-002-23	Sequence 23, Appl
109	34	81.0	1418	2	US-09-010-999-1	Sequence 1, Appl	182	32	76.2	164	2	US-08-854-039B-24	Sequence 24, Appl
110	34	81.0	1418	2	US-09-500-811-20	Sequence 20, Appl	183	32	76.2	164	2	US-08-327-874A-2	Sequence 2, Appl
111	34	81.0	1418	2	US-09-570-573-20	Sequence 20, Appl	184	32	76.2	164	2	US-09-075-505-5	Sequence 5, Appl
112	34	81.0	1418	2	US-09-548-608-20	Sequence 20, Appl	185	32	76.2	164	2	US-08-765-702B-24	Sequence 24, Appl
113	34	81.0	1442	1	US-08-316-650-12	Sequence 12, Appl	186	32	76.2	164	2	US-09-221-268D-3	Sequence 3, Appl
114	34	81.0	1442	4	PCT-US95-02251-12	Sequence 2, Appl	187	32	76.2	164	2	US-09-221-268D-5	Sequence 5, Appl
115	34	81.0	1739	2	US-09-795-061-2	Sequence 2, Appl	188	32	76.2	164	2	US-10-008-960-2	Sequence 2, Appl
116	34	81.0	2149	2	US-09-605-703B-8	Sequence 8, Appl	189	32	76.2	164	4	PCT-US92-10904-2	Sequence 2, Appl
117	33	78.6	19	2	US-09-184-658-20	Sequence 20, Appl	190	32	76.2	164	4	PCT-US94-09700-2	Sequence 2, Appl
118	33	78.6	19	2	US-09-504-262D-20	Sequence 20, Appl	191	32	76.2	164	4	PCT-US94-12936-2	Sequence 2, Appl
119	33	78.6	27	2	US-09-623-548A-1601	Sequence 1601, Ap	192	32	76.2	164	4	PCT-US96-11886A-2	Sequence 2, Appl
120	33	78.6	27	2	US-09-657-276-1601	Sequence 1601, Ap	193	32	76.2	171	2	US-09-011-735-2	Sequence 2, Appl
121	33	78.6	33	1	US-07-972-032-18	Sequence 18, Appl	194	32	76.2	171	2	US-09-029-156-2	Sequence 2, Appl
122	33	78.6	33	1	US-08-477-509B-45	Sequence 45, Appl	195	32	76.2	189	2	US-09-949-016-11225	Sequence 11225, A
123	33	78.6	33	1	US-08-642-255-18	Sequence 18, Appl	196	32	76.2	197	2	US-09-270-767-33275	Sequence 33275, A
124	33	78.6	33	1	US-08-397-633A-9	Sequence 9, Appl	197	32	76.2	197	2	US-09-270-767-48492	Sequence 48492, A
125	33	78.6	33	1	US-08-707-237A-17	Sequence 17, Appl	198	32	76.2	219	2	US-09-252-991A-30360	Sequence 30360, A
126	33	78.6	33	2	US-09-444-791A-45	Sequence 45, Appl	199	32	76.2	221	2	US-09-248-796A-16814	Sequence 16814, A
127	33	78.6	54	1	US-07-972-032-79	Sequence 79, Appl	200	32	76.2	252	2	US-09-270-767-44353	Sequence 44353, A
128	33	78.6	54	1	US-08-642-255-92	Sequence 92, Appl	201	32	76.2	281	2	US-09-252-991A-17748	Sequence 17748, A
129	33	78.6	60	1	US-07-972-032-81	Sequence 81, Appl	202	32	76.2	288	2	US-08-934-494-4	Sequence 4, Appl
130	33	78.6	60	1	US-08-642-255-94	Sequence 94, Appl	203	32	76.2	288	2	US-09-143-068-4	Sequence 4, Appl
131	33	78.6	127	2	US-09-973-378-251	Sequence 251, App	204	32	76.2	288	2	US-09-143-707-4	Sequence 4, Appl
132	33	78.6	128	2	US-09-227-357-190	Sequence 190, App	205	32	76.2	288	2	US-09-202-089-4	Sequence 4, Appl
133	33	78.6	150	2	US-09-636-215-707	Sequence 190, App	206	32	76.2	288	2	US-09-511-133-4	Sequence 4, Appl
134	33	78.6	150	2	US-09-685-166A-707	Sequence 707, App	207	32	76.2	288	2	US-09-690-169-4	Sequence 4, Appl
135	33	78.6	150	2	US-09-679-426-707	Sequence 707, App	208	32	76.2	288	2	US-09-511-631-4	Sequence 4, Appl
136	33	78.6	150	2	US-09-759-143-707	Sequence 707, App	209	32	76.2	288	2	US-09-690-189-4	Sequence 4, Appl
137	33	78.6	150	2	US-09-651-236-707	Sequence 707, App	210	32	76.2	310	2	US-09-219-849-47	Sequence 47, Appl
138	33	78.6	150	2	US-09-657-279-707	Sequence 707, App	211	32	76.2	314	2	US-09-252-991A-20396	Sequence 20396, A
139	33	78.6	150	2	US-10-012-896-707	Sequence 707, App	212	32	76.2	316	2	US-09-252-991A-30877	Sequence 30877, A
140	33	78.6	190	2	US-09-252-991A-18231	Sequence 18231, A	213	32	76.2	342	1	US-08-118-270-9	Sequence 9, Appl
141	33	78.6	203	2	US-09-252-991A-28150	Sequence 28150, A	214	32	76.2	342	1	US-09-252-991A-19222	Sequence 19222, A
142	33	78.6	234	2	US-09-895-674A-1	Sequence 1, Appl	215	32	76.2	342	4	PCT-US93-08528-9	Sequence 9, Appl
143	33	78.6	235	2	US-09-252-991A-30113	Sequence 30113, A	216	32	76.2	351	2	US-09-011-735-1	Sequence 1, Appl
144	33	78.6	305	2	US-08-955-957A-5	Sequence 5, Appl	217	32	76.2	351	2	US-09-029-156-1	Sequence 1, Appl
145	33	78.6	347	2	US-09-623-497-1	Sequence 1, Appl	218	32	76.2	361	2	US-09-369-247-123	Sequence 123, App
146	33	78.6	348	2	US-09-252-991A-21706	Sequence 21706, A	219	32	76.2	361	2	US-10-062-548-123	Sequence 123, App
147	33	78.6	362	2	US-09-252-991A-31641	Sequence 31641, A	220	32	76.2	368	2	US-09-248-796A-16492	Sequence 16492, A
148	33	78.6	468	2	US-09-252-991A-27684	Sequence 27684, A	221	32	76.2	375	2	US-09-600-932-29	Sequence 29, Appl
149	33	78.6	549	1	US-08-494-168-8	Sequence 8, Appl	222	32	76.2	397	2	US-08-327-874A-11	Sequence 11, Appl
150	33	78.6	599	2	US-09-949-016-8890	Sequence 8890, Ap	223	32	76.2	397	2	US-09-252-991A-20113	Sequence 20113, A
151	33	78.6	643	2	US-09-252-991A-17181	Sequence 17181, A	224	32	76.2	397	2	US-10-008-960-11	Sequence 11, Appl
152	33	78.6	676	2	US-09-085-199B-7	Sequence 7, Appl	225	32	76.2	397	4	PCT-US94-09700-11	Sequence 11, Appl
153	33	78.6	722	2	US-09-252-991A-24102	Sequence 24102, A	226	32	76.2	401	2	US-09-219-849-34	Sequence 34, Appl
154	33	78.6	756	2	US-10-104-047-2505	Sequence 2505, Ap	227	32	76.2	401	2	US-09-252-991A-20438	Sequence 20438, A
155	33	78.6	875	2	US-09-252-991A-30056	Sequence 30056, A	228	32	76.2	406	2	US-10-104-047-3291	Sequence 3291, Ap
156	33	78.6	889	2	US-09-489-039A-7241	Sequence 7241, Ap	229	32	76.2	422	2	US-09-252-991A-27513	Sequence 27513, A
157	33	78.6	890	2	US-09-849-602-19	Sequence 19, Appl	230	32	76.2	438	2	US-10-104-047-3166	Sequence 3166, Ap
158	33	78.6	1037	2	US-09-428-711A-21	Sequence 21, Appl	231	32	76.2	440	2	US-09-252-991A-22487	Sequence 22487, A
159	33	78.6	1077	1	US-07-972-032-82	Sequence 82, Appl	232	32	76.2	443	1	US-08-660-765A-4	Sequence 4, Appl
160	33	78.6	1077	1	US-08-642-255-95	Sequence 95, Appl	233	32	76.2	460	2	US-09-826-509-513	Sequence 513, App
161	33	78.6	1712	1	US-09-961-403-9	Sequence 9, Appl	234	32	76.2	479	1	US-08-313-553-7	Sequence 7, Appl
162	32	76.2	20	2	US-08-817-895-13	Sequence 13, Appl	235	32	76.2	479	2	US-08-767-993-7	Sequence 7, Appl
163	32	76.2	20	2	US-08-817-895-16	Sequence 16, Appl	236	32	76.2	492	2	US-08-468-996-12	Sequence 12, Appl
164	32	76.2	20	2	US-09-180-269-9	Sequence 9, Appl	237	32	76.2	509	2	US-09-996-611D-4	Sequence 4, Appl
165	32	76.2	20	2	US-09-180-269-16	Sequence 16, Appl	238	32	76.2	518	1	US-09-252-991A-23604	Sequence 23604, A
166	32	76.2	20	2	US-09-180-269-17	Sequence 17, Appl	239	32	76.2	546	1	US-08-494-168-10	Sequence 10, Appl
167	32	76.2	47	2	US-09-369-247-128	Sequence 128, App	240	32	76.2	593	2	US-09-964-899-15	Sequence 15, Appl
168	32	76.2	47	2	US-10-062-548-128	Sequence 128, App	241	32	76.2	599	2	US-09-602-459-22	Sequence 22, Appl
169	32	76.2	82	2	US-09-621-976-5634	Sequence 5634, Ap	242	32	76.2	599	2	US-09-602-459-23	Sequence 23, Appl
170	32	76.2	102	2	US-09-471-876-1560	Sequence 1560, Ap	243	32	76.2	645	2	US-09-964-899-17	Sequence 17, Appl
171	32	76.2	104	2	US-09-219-849-33	Sequence 33, Appl	244	32	76.2	646	2	US-09-964-899-53	Sequence 53, Appl
172	32	76.2	134	2	US-09-252-991A-29489	Sequence 29489, A	245	32	76.2	728	2	US-09-252-991A-23613	Sequence 23613, A
173	32	76.2	164	1	US-07-970-462A-2	Sequence 2, Appl	246	32	76.2	747	2	US-09-724-864-36	Sequence 36, Appl

247	32	76.2	803	2	US-09-063-035-2	Sequence 2, Appli	320	31	73.8	864	2	US-09-883-096-2	Sequence 2, Appli
248	32	76.2	836	2	US-09-491-356C-21	Sequence 21, Appl	321	31	73.8	979	2	US-08-514-213A-2	Sequence 2, Appli
249	32	76.2	938	2	US-09-949-016-9992	Sequence 992, Ap	322	31	73.8	979	2	US-09-015-399-5	Sequence 5, Appli
250	32	76.2	954	2	US-09-996-611D-1	Sequence 1, Appli	323	31	73.8	1003	2	US-09-949-016-11260	Sequence 17508, A
251	32	76.2	998	2	US-09-252-991A-24402	Sequence 24402, A	324	31	73.8	1064	2	US-09-252-991A-29215	Sequence 29215, A
252	32	76.2	1024	2	US-08-931-820-2	Sequence 2, Appli	325	31	73.8	1113	2	US-09-252-991A-30833	Sequence 30833, A
253	32	76.2	1024	2	US-10-153-469A-30	Sequence 30, Appl	326	31	73.8	1197	2	US-09-252-991A-30833	Sequence 1, Appli
254	32	76.2	1040	2	US-10-153-469A-32	Sequence 32, Appl	327	31	73.8	1971	2	US-09-914-272A-1	Sequence 1, Appli
255	32	76.2	1040	2	US-10-104-889-30	Sequence 30, Appl	328	31	73.8	1971	2	US-10-638-333-1	Sequence 1, Appli
256	32	76.2	1040	2	US-10-104-889-32	Sequence 32, Appl	329	31	73.8	1971	2	US-10-747-133A-1	Sequence 1, Appli
257	32	76.2	1052	1	US-08-852-806-2	Sequence 2, Appli	330	30	71.4	9	2	US-08-159-339A-821	Sequence 821, App
258	32	76.2	1052	1	US-08-852-806-2	Sequence 2, Appli	331	30	71.4	9	2	US-08-159-339A-855	Sequence 855, App
259	32	76.2	1078	2	US-09-163-669-2	Sequence 2, Appli	332	30	71.4	11	2	US-09-586-937-68	Sequence 68, Appl
260	32	76.2	1143	2	US-09-949-016-11185	Sequence 11185, A	333	30	71.4	13	1	US-08-330-599-17	Sequence 17, Appl
261	32	76.2	1298	1	US-08-690-473-2	Sequence 6137, Ap	334	30	71.4	16	2	US-10-226-877A-55	Sequence 55, Appl
262	32	76.2	1298	1	US-08-690-473-2	Sequence 2, Appli	335	30	71.4	18	1	US-07-972-032-5	Sequence 5, Appli
263	32	76.2	1298	2	US-08-843-659-2	Sequence 2, Appli	336	30	71.4	18	1	US-08-642-255-5	Sequence 5, Appli
264	32	76.2	1298	2	US-08-843-659-2	Sequence 2, Appli	337	30	71.4	18	6	5473052-10	Patent No. 5473052
265	32	76.2	1366	2	US-08-963-825-19	Sequence 19, Appl	338	30	71.4	22	2	US-09-219-849-13	Sequence 13, Appl
266	32	76.2	1366	2	US-08-963-825-19	Sequence 19, Appl	339	30	71.4	24	6	5464756-8	Patent No. 5464756
267	32	76.2	1366	2	US-08-570-573-19	Sequence 19, Appl	340	30	71.4	25	1	US-08-383-348A-6	Sequence 6, Appli
268	32	76.2	1366	2	US-09-548-608-19	Sequence 19, Appl	341	30	71.4	25	1	US-08-383-348A-7	Sequence 7, Appli
269	32	76.2	1366	2	US-09-585-887-10	Sequence 10, Appl	342	30	71.4	25	1	US-08-383-348A-8	Sequence 8, Appli
270	32	76.2	1366	2	US-09-585-887-10	Sequence 10, Appl	343	30	71.4	25	1	US-08-383-348A-9	Sequence 9, Appli
271	32	76.2	1366	2	US-09-949-016-5882	Sequence 5882, Ap	344	30	71.4	25	1	US-08-383-348A-10	Sequence 10, Appl
272	32	76.2	1694	1	US-08-494-168-2	Sequence 2, Appli	345	30	71.4	25	1	US-08-383-348A-11	Sequence 11, Appl
273	32	76.2	1958	1	US-07-945-283-2	Sequence 2, Appli	346	30	71.4	25	1	US-08-383-348A-12	Sequence 12, Appl
274	31	73.8	9	2	US-08-159-339A-822	Sequence 822, App	347	30	71.4	25	1	US-08-383-348A-13	Sequence 13, Appl
275	31	73.8	148	2	US-09-252-991A-31175	Sequence 31175, A	348	30	71.4	25	1	US-08-383-348A-14	Sequence 14, Appl
276	31	73.8	149	2	US-09-252-991A-20244	Sequence 20244, A	349	30	71.4	25	1	US-08-383-804B-6	Sequence 6, Appli
277	31	73.8	149	2	US-09-311-021-148	Sequence 148, App	350	30	71.4	25	1	US-08-383-804B-7	Sequence 7, Appli
278	31	73.8	167	2	US-09-252-991A-22294	Sequence 22294, A	351	30	71.4	25	1	US-08-383-804B-8	Sequence 8, Appli
279	31	73.8	227	2	US-09-252-991A-21816	Sequence 21816, A	352	30	71.4	25	1	US-08-383-804B-9	Sequence 9, Appli
280	31	73.8	228	2	US-09-219-849-38	Sequence 38, Appl	353	30	71.4	25	1	US-08-383-804B-10	Sequence 10, Appl
281	31	73.8	239	2	US-09-252-991A-32898	Sequence 32898, A	354	30	71.4	25	1	US-08-383-804B-11	Sequence 11, Appl
282	31	73.8	261	2	US-09-902-540-14918	Sequence 14918, A	355	30	71.4	25	1	US-08-383-804B-12	Sequence 12, Appl
283	31	73.8	261	2	US-09-252-991A-29810	Sequence 29810, A	356	30	71.4	25	1	US-08-383-804B-13	Sequence 13, Appl
284	31	73.8	285	2	US-10-104-047-3028	Sequence 3028, Ap	357	30	71.4	25	1	US-08-383-804B-14	Sequence 14, Appl
285	31	73.8	320	2	US-09-949-016-6177	Sequence 6177, Ap	358	30	71.4	25	1	US-08-383-748A-6	Sequence 6, Appli
286	31	73.8	338	2	US-09-883-096-4	Sequence 10451, A	359	30	71.4	25	1	US-08-383-748A-7	Sequence 7, Appli
287	31	73.8	338	2	US-09-883-096-4	Sequence 4, Appli	360	30	71.4	25	1	US-08-383-748A-8	Sequence 8, Appli
288	31	73.8	356	2	US-09-902-540-12881	Sequence 12881, A	361	30	71.4	25	1	US-08-383-748A-9	Sequence 9, Appli
289	31	73.8	372	2	US-09-252-991A-30132	Sequence 30132, A	362	30	71.4	25	1	US-08-383-748A-10	Sequence 10, Appl
290	31	73.8	383	2	US-09-883-096-5	Sequence 5, Appli	363	30	71.4	25	1	US-08-383-748A-11	Sequence 11, Appl
291	31	73.8	388	2	US-09-252-991A-30804	Sequence 30804, A	364	30	71.4	25	1	US-08-383-748A-12	Sequence 12, Appl
292	31	73.8	394	2	US-09-252-991A-26156	Sequence 26156, A	365	30	71.4	25	1	US-08-383-748A-13	Sequence 13, Appl
293	31	73.8	400	2	US-09-489-039A-7483	Sequence 7483, Ap	366	30	71.4	25	1	US-08-383-748A-14	Sequence 14, Appl
294	31	73.8	402	2	US-09-252-991A-16708	Sequence 16708, A	367	30	71.4	25	1	US-08-383-748A-14	Sequence 6, Appli
295	31	73.8	407	2	US-09-252-991A-21200	Sequence 21200, A	368	30	71.4	25	1	US-08-814-309A-7	Sequence 7, Appli
296	31	73.8	413	2	US-09-902-540-15895	Sequence 15895, A	369	30	71.4	25	1	US-08-814-309A-8	Sequence 8, Appli
297	31	73.8	489	1	US-08-794-785-7	Sequence 7, Appli	370	30	71.4	25	1	US-08-814-309A-9	Sequence 9, Appli
298	31	73.8	489	2	US-09-249-200-7	Sequence 7, Appli	371	30	71.4	25	1	US-08-814-309A-10	Sequence 10, Appl
299	31	73.8	493	2	US-09-252-991A-16925	Sequence 16925, A	372	30	71.4	25	1	US-08-814-309A-11	Sequence 11, Appl
300	31	73.8	516	2	US-09-252-991A-21880	Sequence 21880, A	373	30	71.4	25	1	US-08-814-309A-12	Sequence 12, Appl
301	31	73.8	517	2	US-09-252-991A-32085	Sequence 32085, A	374	30	71.4	25	1	US-08-814-309A-13	Sequence 13, Appl
302	31	73.8	518	1	US-08-392-367B-2	Sequence 2, Appli	375	30	71.4	25	1	US-08-814-309A-14	Sequence 14, Appl
303	31	73.8	518	2	US-08-893-467A-2	Sequence 2, Appli	376	30	71.4	44	2	US-09-419-381-422	Sequence 122, App
304	31	73.8	519	2	US-09-252-991A-26230	Sequence 26230, A	377	30	71.4	52	2	US-09-205-258-960	Sequence 960, App
305	31	73.8	520	2	US-09-252-991A-32505	Sequence 32505, A	378	30	71.4	52	2	US-10-004-860-960	Sequence 960, App
306	31	73.8	529	2	US-09-902-540-16796	Sequence 16796, A	379	30	71.4	96	2	US-09-621-976-4599	Sequence 4599, Ap
307	31	73.8	532	1	US-08-494-168-9	Sequence 9, Appli	380	30	71.4	124	2	US-09-252-991A-20762	Sequence 20762, A
308	31	73.8	557	2	US-09-320-095-10	Sequence 10, Appl	381	30	71.4	129	2	US-09-883-777-5	Sequence 5, Appli
309	31	73.8	557	2	US-09-523-487-10	Sequence 10, Appl	382	30	71.4	129	2	US-09-743-454A-5	Sequence 5, Appli
310	31	73.8	595	2	US-09-252-991A-30780	Sequence 30780, A	383	30	71.4	136	2	US-09-252-991A-25007	Sequence 25007, A
311	31	73.8	623	1	US-08-653-740-7	Sequence 7, Appli	384	30	71.4	146	2	US-09-252-991A-32438	Sequence 32438, A
312	31	73.8	623	1	US-09-073-594-7	Sequence 7, Appli	385	30	71.4	148	2	US-09-543-681A-6030	Sequence 6030, Ap
313	31	73.8	623	2	US-09-275-925-7	Sequence 7, Appli	386	30	71.4	171	2	US-09-252-991A-32311	Sequence 32311, A
314	31	73.8	625	2	US-09-252-991A-19871	Sequence 19871, A	387	30	71.4	182	2	US-09-640-211A-863	Sequence 863, App
315	31	73.8	637	2	US-09-489-039A-13362	Sequence 13362, A	388	30	71.4	183	2	US-09-489-039A-7289	Sequence 7289, Ap
316	31	73.8	638	2	US-10-001-887-108	Sequence 108, App	389	30	71.4	194	2	US-09-252-991A-21143	Sequence 21143, A
317	31	73.8	716	2	US-09-252-991A-30683	Sequence 30683, A	390	30	71.4	195	2	US-09-252-991A-30084	Sequence 30084, A
318	31	73.8	783	2	US-09-252-991A-19698	Sequence 19698, A	391	30	71.4	197	2	US-09-834-759-516	Sequence 516, App
319	31	73.8	825	2	US-09-540-824-26	Sequence 26, Appl	392	30	71.4	197	2	US-10-076-622-516	Sequence 516, App

393	30	71.4	201	2	US-09-902-540-14396	Sequence 14396, A	466	30	71.4	276	2	US-09-205-258-958	Sequence 958, App
394	30	71.4	203	2	US-08-855-825-12	Sequence 12, Appl	467	30	71.4	276	2	US-10-004-860-958	Sequence 958, App
395	30	71.4	203	2	US-08-855-825-14	Sequence 14, Appl	468	30	71.4	277	2	US-09-892-081-5	Sequence 5, Appl
396	30	71.4	204	2	US-08-815-783-6	Sequence 6, Appl	469	30	71.4	278	2	US-09-834-759-515	Sequence 515, App
397	30	71.4	205	1	US-09-912-227-4	Sequence 4, Appl	470	30	71.4	278	2	US-10-076-622-515	Sequence 515, App
398	30	71.4	205	2	US-08-883-086-8	Sequence 8, Appl	471	30	71.4	279	2	US-09-010-999-2	Sequence 2, Appl
399	30	71.4	205	2	US-09-589-287B-4	Sequence 4, Appl	472	30	71.4	280	2	US-09-247-155-178	Sequence 178, App
400	30	71.4	205	2	US-09-513-584-4	Sequence 4, Appl	473	30	71.4	280	2	US-09-903-190-178	Sequence 178, App
401	30	71.4	205	2	US-09-027-287-5	Sequence 4, Appl	474	30	71.4	283	2	US-09-252-991A-22696	Sequence 22696, A
402	30	71.4	205	2	US-09-027-287-5	Sequence 5, Appl	475	30	71.4	287	2	US-09-252-991A-32232	Sequence 32232, A
403	30	71.4	205	2	US-09-252-656B-4	Sequence 4, Appl	476	30	71.4	291	1	US-08-284-941-9	Sequence 9, Appl
404	30	71.4	205	2	US-09-252-656B-5	Sequence 5, Appl	477	30	71.4	291	1	US-08-447-642-9	Sequence 9, Appl
405	30	71.4	205	2	US-10-082-260-6	Sequence 6, Appl	478	30	71.4	291	2	US-09-236-503-9	Sequence 9, Appl
406	30	71.4	205	2	US-09-345-790-4	Sequence 4, Appl	479	30	71.4	291	4	PCT-US93-02147A-9	Sequence 9, Appl
407	30	71.4	205	2	US-09-879-919-6	Sequence 6, Appl	480	30	71.4	309	2	US-09-724-623-72	Sequence 72, Appl
408	30	71.4	205	2	US-09-588-947A-4	Sequence 4, Appl	481	30	71.4	324	2	US-09-252-991A-29703	Sequence 29703, A
409	30	71.4	205	2	US-09-131-237C-4	Sequence 4, Appl	482	30	71.4	330	2	US-09-270-767-41611	Sequence 41611, A
410	30	71.4	205	2	US-09-589-286A-4	Sequence 4, Appl	483	30	71.4	376	2	US-10-094-944-28	Sequence 28, Appl
411	30	71.4	205	2	US-09-523-323-4	Sequence 4, Appl	484	30	71.4	387	2	US-09-252-991A-21811	Sequence 21811, A
412	30	71.4	205	2	US-09-523-323-5	Sequence 5, Appl	485	30	71.4	391	2	US-09-252-991A-25588	Sequence 25588, A
413	30	71.4	205	2	US-09-005-874-4	Sequence 4, Appl	486	30	71.4	402	2	US-09-252-991A-20683	Sequence 20683, A
414	30	71.4	205	2	US-09-255-794A-4	Sequence 4, Appl	487	30	71.4	407	2	US-10-104-047-3895	Sequence 3895, App
415	30	71.4	205	2	US-09-507-968D-4	Sequence 4, Appl	488	30	71.4	415	2	US-09-032-523-2	Sequence 2, Appl
416	30	71.4	205	2	US-09-246-129B-4	Sequence 4, Appl	489	30	71.4	415	2	US-09-802-633-2	Sequence 2, Appl
417	30	71.4	205	2	US-09-589-285-4	Sequence 4, Appl	490	30	71.4	415	2	US-09-907-794A-104	Sequence 104, App
418	30	71.4	210	2	US-09-949-002-304	Sequence 304, App	491	30	71.4	415	2	US-09-905-125A-104	Sequence 104, App
419	30	71.4	210	2	US-09-489-039A-9923	Sequence 9923, App	492	30	71.4	415	2	US-09-302-775A-104	Sequence 104, App
420	30	71.4	229	2	US-09-800-729-117	Sequence 117, App	493	30	71.4	415	2	US-09-906-700-104	Sequence 104, App
421	30	71.4	231	2	US-09-252-991A-22663	Sequence 22663, A	494	30	71.4	415	2	US-09-903-603A-104	Sequence 104, App
422	30	71.4	232	2	US-09-834-759-517	Sequence 517, App	495	30	71.4	415	2	US-09-904-920A-104	Sequence 104, App
423	30	71.4	232	2	US-10-076-622-517	Sequence 517, App	496	30	71.4	415	2	US-09-909-064-104	Sequence 104, App
424	30	71.4	237	2	US-09-252-991A-19656	Sequence 19656, A	497	30	71.4	415	2	US-09-905-381A-104	Sequence 104, App
425	30	71.4	240	2	US-09-538-092-930	Sequence 930, App	498	30	71.4	415	2	US-09-906-618-104	Sequence 104, App
426	30	71.4	243	2	US-09-489-847-205	Sequence 205, App	499	30	71.4	415	2	US-09-906-646-104	Sequence 104, App
427	30	71.4	243	2	US-09-692-081-4	Sequence 4, Appl	500	30	71.4	415	2	US-09-904-462-104	Sequence 104, App
428	30	71.4	243	2	US-09-834-759-514	Sequence 514, App							
429	30	71.4	243	2	US-10-012-231A-352	Sequence 352, App							
430	30	71.4	243	2	US-10-012-231A-431	Sequence 431, App							
431	30	71.4	243	2	US-10-015-389A-352	Sequence 352, App							
432	30	71.4	243	2	US-10-015-389A-431	Sequence 431, App							
433	30	71.4	243	2	US-10-006-768A-352	Sequence 352, App							
434	30	71.4	243	2	US-10-006-768A-431	Sequence 431, App							
435	30	71.4	243	2	US-10-015-671A-352	Sequence 352, App							
436	30	71.4	243	2	US-10-015-671A-431	Sequence 431, App							
437	30	71.4	243	2	US-10-015-393A-352	Sequence 352, App							
438	30	71.4	243	2	US-10-015-393A-431	Sequence 431, App							
439	30	71.4	243	2	US-10-011-833A-352	Sequence 352, App							
440	30	71.4	243	2	US-10-011-833A-431	Sequence 431, App							
441	30	71.4	243	2	US-10-006-041A-352	Sequence 352, App							
442	30	71.4	243	2	US-10-006-041A-431	Sequence 431, App							
443	30	71.4	243	2	US-10-012-064A-352	Sequence 352, App							
444	30	71.4	243	2	US-10-012-064A-431	Sequence 431, App							
445	30	71.4	243	2	US-09-252-991A-22242	Sequence 22242, A							
446	30	71.4	243	2	US-09-800-729-85	Sequence 85, Appl							
447	30	71.4	245	2	US-09-800-729-115	Sequence 115, App							
448	30	71.4	245	2	US-09-800-729-116	Sequence 116, App							
449	30	71.4	245	2	US-09-800-729-118	Sequence 118, App							
450	30	71.4	245	2	US-09-800-729-119	Sequence 119, App							
451	30	71.4	245	2	US-09-692-081-2	Sequence 2, Appl							
452	30	71.4	245	2	US-09-311-021-104	Sequence 104, App							
453	30	71.4	248	2	US-09-949-016-10554	Sequence 10554, A							
454	30	71.4	249	2	US-09-917-254-96	Sequence 96, Appl							
455	30	71.4	249	2	US-09-902-540-10521	Sequence 10521, A							
456	30	71.4	251	2	US-09-949-002-541	Sequence 541, App							
457	30	71.4	255	2	US-09-883-134-11	Sequence 11, Appl							
458	30	71.4	255	2	US-10-306-879-11	Sequence 11, Appl							
459	30	71.4	259	2	US-09-198-452A-146	Sequence 146, App							
460	30	71.4	264	2	US-09-438-185A-130	Sequence 130, App							
461	30	71.4											
462	30	71.4											
463	30	71.4											
464	30	71.4											
465	30	71.4											

ALIGNMENTS

RESULT 1

US-09-252-991A-31949

Sequence 31949, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 31949

LENGTH: 495

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31949

Query Match

Best Local Similarity

Matches

Score 39;

DB 2;

Length 495;

87.5%;

Pred. No. 95;

Conservative

1;

Mismatches

0;

Indels

0;

Gaps

0;

QY

1

GREGSPGL

8

Db

208

GREGSPGV

215

RESULT 2

US-10-153-469A-44

; Sequence 44, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/153.469A
; FILING DATE: 22-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-153-469A-44
Query Match 90.5%; Score 38; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26
RESULT 3
US-10-153-469A-46
; Sequence 46, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/10/153.469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-153-469A-46
Query Match 90.5%; Score 38; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26
RESULT 4
US-10-104-889-44
; Sequence 44, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-104-889-44

Query Match 90.5%; Score 38; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 20 GAEGSPG 26

RESULT 5

US-10-104-889-46
; Sequence 46, Application US/10104889
; Patent No. 6958223

; GENERAL INFORMATION:

; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI

; PAOLELLA, DAVID

; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DILWORTH & BARRESE

; STREET: 333 EARLE OVINGTON BOULEVARD

; CITY: UNIONDALE

; STATE: NY

; COUNTRY: U.S.A.

; ZIP: 11553

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/104,889

; FILING DATE: 22-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/169,768

; FILING DATE: 09-OCT-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: STEEN, JEFFREY S

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (516) 228-8484

; TELEFAX: (516) 228-8516

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 219 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-104-889-46

Query Match 90.5%; Score 38; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 20 GAEGSPG 26

, RESULT 6

US-09-252-991A-16703
; Sequence 16703, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16703

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16703

Query Match 90.5%; Score 38; DB 2; Length 238;

Best Local Similarity 100.0%; Pred. No. 69;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

| | | | |

Db 56 GAEGSPG 62

RESULT 7

US-09-252-991A-32551

; Sequence 32551, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 32551

; LENGTH: 926

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32551

Query Match 90.5%; Score 38; DB 2; Length 926;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

| | | | |

Db 875 GAEGSPG 881

RESULT 8

US-08-931-820-1

; Sequence 1, Application US/08931820

; Patent No. 6010863

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Assay for collagen degradation

; NUMBER OF SEQUENCES: 4

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/931,820
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: EP 96202596.1
;; FILING DATE:
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1057 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 90.5%; Score 38; DB 2; Length 1057;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 9
US-10-153-469A-20
; Sequence 20, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,469A
; FILING DATE: 22-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown

;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-153-469A-20

Query Match 90.5%; Score 38; DB 2; Length 1057;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 10
US-10-104-889-20
; Sequence 20, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown

;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-104-889-20

Query Match 90.5%; Score 38; DB 2; Length 1057;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 11
US-10-153-469A-11
; Sequence 11, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-153-469A-11

Query Match 90.5%; Score 38; DB 2; Length 1107;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 12
US-10-104-889-11
Sequence 11, Application US/10104889
Patent No. 6958223
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1107 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-889-11

Query Match 90.5%; Score 38; DB 2; Length 1107;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 13
US-10-153-469A-8
Sequence 8, Application US/10153469A
Patent No. 6927287
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1171 amino acids
TYPE: amino acid
STRANDEDNESS: single


```
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-09-500-811-18

Query Match 90.5%; Score 38; DB 2; Length 1341;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 895 GAEGSPG 901

RESULT 17
US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Oviat, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
```

```
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-09-570-573-18

Query Match 90.5%; Score 38; DB 2; Length 1341;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 895 GAEGSPG 901

RESULT 18
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Oviat, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18

Query Match 90.5%; Score 38; DB 2; Length 1341;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 895 GAEGSPG 901

RESULT 19

US-10-153-469A-10
; Sequence 10, Application US/10153469A

; Patent No. 6927287
; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.

BUECHTER, DOUGLAS

BROKAW, JANE

ZHANG, GUANGHUI

PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,469A

FILING DATE: 22-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION NUMBER: US/09/169,768

FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1388 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-153-469A-10

Query Match 90.5%; Score 38; DB 2; Length 1388;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 858 GAEGSPG 864

RESULT 20

US-10-104-889-10

; Sequence 10, Application US/10104889

; Patent No. 6958223

; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.

BUECHTER, DOUGLAS

BROKAW, JANE

ZHANG, GUANGHUI

PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889

FILING DATE: 22-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768

FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1388 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 10:

US-10-104-889-10

Query Match 90.5%; Score 38; DB 2; Length 1388;

Best Local Similarity 100.0%; Pred. No. 3.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
| | | | |
Db 858 GAEGSPG 864

RESULT 21

US-09-585-887-9

; Sequence 9, Application US/09585887

; Patent No. 6413742

; GENERAL INFORMATION:

APPLICANT: Olsen, David R

APPLICANT: Chang, Robert

APPLICANT: McMullin, Hugh

APPLICANT: Hitzeman, Ronald A.

APPLICANT: Chisholm, George

TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND

TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT

TITLE OF INVENTION: CELLS

FILE REFERENCE: 225002030400

CURRENT APPLICATION NUMBER: US/09/585,887

CURRENT FILING DATE: 2000-05-31

PRIOR APPLICATION NUMBER: 09/289,578

PRIOR FILING DATE: 1999-04-09

PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

Query Match 90.5%; Score 38; DB 2; Length 1461;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 1016 GAEGSPG 1022

RESULT 22

US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978

; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.

; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS

; FILE REFERENCE: 225002030400

; CURRENT APPLICATION NUMBER: US/09/289,578

; CURRENT FILING DATE: 1999-04-10

; PRIOR APPLICATION NUMBER: 60/084,828

; PRIOR FILING DATE: 1998-05-08

; NUMBER OF SEQ ID NOS: 11

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1461

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-289-578-9

Query Match 90.5%; Score 38; DB 2; Length 1461;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 1016 GAEGSPG 1022

RESULT 23

US-09-331-347C-21

; Sequence 21, Application US/09331347C

; Patent No. 6617431

; GENERAL INFORMATION:

; APPLICANT: Meristem Therapeutics, S.A.

; APPLICANT: Meristem Therapeutics, S.A.

; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me

; TITLE OF INVENTION: Obtaining Such and Their Uses

; FILE REFERENCE: 1149-3

; CURRENT APPLICATION NUMBER: US/09/331,347C

; CURRENT FILING DATE: 1999-08-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 21

; LENGTH: 1464

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-331-347C-21

Query Match 90.5%; Score 38; DB 2; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 1019 GAEGSPG 1025

RESULT 24

US-10-360-101-221

; Sequence 221, Application US/10360101

; Patent No. 6861236

; GENERAL INFORMATION:

; APPLICANT: Moll, Gert N.

; APPLICANT: Leenhouts, Cornelis J.

; TITLE OF INVENTION: Export and modification of (poly)peptide in the lantibiotic way

; FILE REFERENCE: 2183-5673

; CURRENT APPLICATION NUMBER: US/10/360,101

; CURRENT FILING DATE: 2003-02-07

; PRIOR APPLICATION NUMBER: EP 02077060.8

; PRIOR FILING DATE: 2002-05-24

; NUMBER OF SEQ ID NOS: 309

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 221

; LENGTH: 1027

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: sequence of collagen

; NAME/KEY: MISC FEATURE

; LOCATION: (96)..(937)

; OTHER INFORMATION: No. 6861236e = No. 6861236e = "Xaa" at positions 96 through 937 me

US-10-360-101-221

Query Match 88.1%; Score 37; DB 2; Length 1027;
Best Local Similarity 87.5%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 739 GAEGSPGL 746

RESULT 25

US-09-949-016-7065

; Sequence 7065, Application US/09949016

; Patent No. 6812339

; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED

; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14

; PRIOR APPLICATION NUMBER: 60/241,755

; PRIOR FILING DATE: 2000-10-20

; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7065

; LENGTH: 1218

; TYPE: PRT

; ORGANISM: Human

US-09-949-016-7065

Query Match 88.1%; Score 37; DB 2; Length 1218;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Oy 1 GAEGSPGL 8
Db 64 GAEGPPGL 71

RESULT 26
US-09-949-016-5884
; Sequence 5884, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5884
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-5884

Query Match 88.1%; Score 37; DB 2; Length 1690;
Best Local Similarity 87.5%; Pred. No. 6.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GAEGSPGL 8
Db 533 GAEGPPGL 540

RESULT 27
US-09-902-540-12854
; Sequence 12854, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 12854
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-12854

Query Match 85.7%; Score 36; DB 2; Length 376;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AEGSPGL 8
Db 184 AEGPPGL 190

RESULT 28
US-09-066-046-8

; Sequence 8, Application US/09066046A
; Patent No. 6204252
; GENERAL INFORMATION:
; APPLICANT: STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC
; EHRlichia AND METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,046A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106.941.155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 484 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-066-046-8

Query Match 85.7%; Score 36; DB 2; Length 484;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 AEGSPGL 8
Db 15 AEGPPGL 21

RESULT 29
US-09-538-092-1269
; Sequence 1269, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 1269
; LENGTH: 486
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q05215
US-09-538-092-1269

Query Match      85.7%; Score 36; DB 2; Length 486;
Best Local Similarity 87.5%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      |||||
Db      279 GAEGLPGL 286

RESULT 30
US-09-949-016-6151
; Sequence 6151, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6151
; LENGTH: 486
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6151

Query Match      85.7%; Score 36; DB 2; Length 486;
Best Local Similarity 87.5%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      |||||
Db      279 GAEGLPGL 286

RESULT 31
US-09-949-016-10733
; Sequence 10733, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10733
; LENGTH: 510
; TYPE: PRT
```

```
; ORGANISM: Human
US-09-949-016-10733

Query Match      85.7%; Score 36; DB 2; Length 510;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      |||||
Db      303 GAEGLPGL 310

RESULT 32
US-09-066-046-2
; Sequence 2, Application US/09066046A
; Patent No. 6204252
; GENERAL INFORMATION:
; APPLICANT: MURPHY, Cheryl
; STOREY, James
; BELTZ, Gerald A.
; COUGHLIN, Richard T.
; TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTTIC
; EHRUCHIA AND METHODS OF USE
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSES: HALE AND DORR LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: United States
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,046A
; FILING DATE: 24-Apr-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Superko, Colleen
; REGISTRATION NUMBER: 39,850
; REFERENCE/DOCKET NUMBER: 106.941.155
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 526-6000
; TELEFAX: (617) 526-5000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 619 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-066-046-2

Query Match      85.7%; Score 36; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 AEGSPGL 8
      |||||
Db      15 AEGSPGL 21

RESULT 33
US-09-949-016-6136
; Sequence 6136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
```

```
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT FILING DATE: 2000-04-14
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6136
; LENGTH: 1609
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6136

Query Match      85.7%; Score 36; DB 2; Length 1603;
Best Local Similarity 75.0%; Pred. No. 8.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      820 GAQGSFV 827

RESULT 34
US-09-949-016-10910
; Sequence 10910, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10910
; LENGTH: 1609
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10910

Query Match      85.7%; Score 36; DB 2; Length 1609;
Best Local Similarity 75.0%; Pred. No. 8.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      826 GAQGSFV 833

RESULT 35
US-09-795-061-4
; Sequence 4, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Imamura, Yasutada
; TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
```

```
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-061-4

Query Match      85.7%; Score 36; DB 2; Length 1745;
Best Local Similarity 75.0%; Pred. No. 9.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      1184 GSEGTPL 1191

RESULT 36
US-09-949-002-405
; Sequence 405, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-405

Query Match      85.7%; Score 36; DB 2; Length 1745;
Best Local Similarity 75.0%; Pred. No. 9.6e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      1184 GSEGTPL 1191

RESULT 37
US-09-949-002-492
; Sequence 492, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 492
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-492

Query Match      85.7%; Score 36; DB 2; Length 1771;
Best Local Similarity 75.0%; Pred. No. 9.7e+02;
```


Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|:|:|:|
Db 1210 GSEGTPL 1217

RESULT 38

US-09-919-497-56
; Sequence 56, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56

Query Match 85.7%; Score 36; DB 2; Length 1806;
Best Local Similarity 75.0%; Pred. No. 9.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
|:|:|:|
Db 1036 GAQAPGL 1043

RESULT 39

US-07-609-716-61
; Sequence 61, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-61

Query Match 83.3%; Score 35; DB 1; Length 62;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|:|:|:|
Db 46 GADGSPG 52

RESULT 40

US-08-642-255-31
; Sequence 31, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-642-255-31

Query Match 83.3%; Score 35; DB 1; Length 62;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|:|:|:|
Db 46 GADGSPG 52

RESULT 41
US-08-475-411A-61
; Sequence 61: Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-475-411A-61
Query Match 83.3%; Score 35; DB 2; Length 62;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GAEGSPG 7
|||
Db 46 GADGSPG 52
RESULT 42
US-08-478-029A-61
; Sequence 61: Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-478-029A-61
Query Match 83.3%; Score 35; DB 2; Length 62;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Oy 1 GAEGSPG 7
|||
Db 46 GADGSPG 52
RESULT 43
US-07-609-716-111
; Sequence 111: Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA: US/07/609,716
/ APPLICATION NUMBER: US/07/609,716
/ FILING DATE: 06-NOV-1990
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rowland, Bertram I
/ REGISTRATION NUMBER: 20015
/ REFERENCE/DOCKET NUMBER: A-55186-3/BIR
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-781-1989
/ TELEFAX: 415-398-3249
/ INFORMATION FOR SEQ ID NO: 111:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 69 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-07-609-716-111

Query Match 83.3%; Score 35; DB 1; Length 69;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
||:||||
Db 43 GADGSPG 49

RESULT 44
US-08-475-411A-111
/ Sequence 111, Application US/08475411A
/ Patent No. 6140072
/ GENERAL INFORMATION:
/ APPLICANT: Ferrari, Franco A.
/ APPLICANT: Cappello, Joseph
/ TITLE OF INVENTION: Functional Recombinantly Prepared
/ TITLE OF INVENTION: Synthetic Protein Polymer
/ NUMBER OF SEQUENCES: 119
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
/ STREET: Four Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: US
/ ZIP: 94111
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/475,411A
/ FILING DATE: 07-JUN-1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/609,716
/ FILING DATE: 06-NOV-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/269,429
/ FILING DATE: 09-NOV-1988
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/114,618
/ FILING DATE: 04-NOV-1986
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trecartin, Richard F.
/ REGISTRATION NUMBER: 31,801
/ REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK

/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-781-1989
/ TELEFAX: 415-398-3249
/ INFORMATION FOR SEQ ID NO: 111:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 69 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-475-411A-111

Query Match 83.3%; Score 35; DB 2; Length 69;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
||:||||
Db 43 GADGSPG 49

RESULT 45
US-08-478-029A-111
/ Sequence 111, Application US/08478029A
/ Patent No. 6184348
/ GENERAL INFORMATION:
/ APPLICANT: Ferrari, Franco A.
/ APPLICANT: Cappello, Joseph
/ TITLE OF INVENTION: Functional Recombinantly Prepared
/ TITLE OF INVENTION: Synthetic Protein Polymer
/ NUMBER OF SEQUENCES: 119
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
/ STREET: Four Embarcadero Center, Suite 3400
/ CITY: San Francisco
/ STATE: CA
/ COUNTRY: US
/ ZIP: 94111
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/478,029A
/ FILING DATE: 07-JUN-1995
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/609,716
/ FILING DATE: 06-NOV-1990
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/269,429
/ FILING DATE: 09-NOV-1988
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/114,618
/ FILING DATE: 29-OCT-1987
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 06/927,258
/ FILING DATE: 04-NOV-1986
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Trecartin, Richard F.
/ REGISTRATION NUMBER: 31,801
/ REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 415-781-1989
/ TELEFAX: 415-398-3249
/ INFORMATION FOR SEQ ID NO: 111:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 69 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide

US-08-478-029A-111

Query Match 83.3%; Score 35; DB 2; Length 69;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGSPG 7
||:||||
Db 43 GADGSPG 49

RESULT 46

US-07-609-716-113
; Sequence 113, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-609-716-113

Query Match 83.3%; Score 35; DB 1; Length 72;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGSPG 7
||:||||
Db 43 GADGSPG 49

RESULT 47

US-08-475-411A-113
; Sequence 113, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RPT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249

INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-475-411A-113

Query Match 83.3%; Score 35; DB 2; Length 72;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGSPG 7
||:||||
Db 43 GADGSPG 49

RESULT 48

US-08-478-029A-113
; Sequence 113, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,029A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 72 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-478-029A-113

Query Match 83.3%; Score 35; DB 2; Length 72;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 43 GADGSPG 49

RESULT 49
US-07-609-716-114
Sequence 114, Application US/07609716
Patent No. 5514581
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 82 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-609-716-114

Query Match 83.3%; Score 35; DB 1; Length 82;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 43 GADGSPG 49

RESULT 50
US-08-475-411A-114
Sequence 114, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,411A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-411A-114

Query Match 83.3%; Score 35; DB 2; Length 85;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
||:||||
Db 46 GADGSPG 52

RESULT 51

US-08-478-029A-114
; Sequence 114, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 114:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-478-029A-114

Query Match 83.3%; Score 35; DB 2; Length 85;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
||:||||
Db 46 GADGSPG 52

RESULT 52

US-09-902-540-10883
; Sequence 10883, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10883
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; US-09-902-540-10883

Query Match 83.3%; Score 35; DB 2; Length 115;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||:||||
Db 40 GAGTGPGL 47

RESULT 53

US-09-949-016-9265
; Sequence 9265, Application US/09949016
; Patent No. 8812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9265
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-9265

Query Match 83.3%; Score 35; DB 2; Length 313;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
||:||||
Db 200 GAGQDPGL 207

RESULT 54

US-07-609-716-66
; Sequence 66, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared

; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 118
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/07/609,716
; FILING DATE: 06-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-609-716-66

Query Match 83.3%; Score 35; DB 1; Length 357;
Best Local Similarity 85.7%; Pred. No. 3.le+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 79 GADGSPG 85
||:||||

RESULT 55
US-08-642-255-33
; Sequence 33, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHNACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-33

Query Match 83.1%; Score 35; DB 1; Length 357;
Best Local Similarity 85.7%; Pred. No. 3.le+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 79 GADGSPG 85
||:||||

RESULT 56
US-08-475-411A-66
; Sequence 66, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-475-411A-66

Query Match 83.3%; Score 35; DB 2; Length 357;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 79 GADGSPG 85

RESULT 57

US-08-478-029A-66
; Sequence 66, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-478-029A-66

Query Match 83.3%; Score 35; DB 2; Length 357;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 79 GADGSPG 85

RESULT 58

US-09-252-991A-20479
; Sequence 20479, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20479
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20479

Query Match 83.3%; Score 35; DB 2; Length 377;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 253 GAEGTPG 259

RESULT 59

US-09-902-540-13248
; Sequence 13248, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 13248
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-13248

Query Match 83.3%; Score 35; DB 2; Length 430;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 398 GAEGAPG 404

RESULT 60

US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.

```
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; PREPARATION THEREOF
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match      83.3%; Score 35; DB 2; Length 595;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      573 GADGSPG 579

RESULT 61
US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; PREPARATION THEREOF
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-50

Query Match      83.3%; Score 35; DB 2; Length 595;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      573 GADGSPG 579

RESULT 62
US-09-219-348-3
; Sequence 3, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2

Query Match      83.3%; Score 35; DB 2; Length 623;
Best Local Similarity 85.7%; Pred. No. 5.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      333 GSEGPSG 339

RESULT 63
US-09-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2

Query Match      83.3%; Score 35; DB 2; Length 626;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      333 GSEGPSG 339

RESULT 64
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; PREPARATION THEREOF
US-09-219-849-50

Query Match      83.3%; Score 35; DB 2; Length 595;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      573 GADGSPG 579
```

```
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-3

Query Match      83.3%; Score 35; DB 2; Length 623;
Best Local Similarity 85.7%; Pred. No. 5.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      333 GSEGPSG 339

RESULT 63
US-09-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2

Query Match      83.3%; Score 35; DB 2; Length 626;
Best Local Similarity 85.7%; Pred. No. 5.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      333 GSEGPSG 339

RESULT 64
US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; PREPARATION THEREOF
US-09-219-849-50
```

```
;
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match      83.3%; Score 35; DB 2; Length 822;
Best Local Similarity 85.7%; Pred. No. 6.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      573 GADGSPG 579

RESULT 65
US-10-210-428-1
; Sequence 1, Application US/10210428
; Patent No. 6814969
; GENERAL INFORMATION:
; APPLICANT: David M. Koelle
; APPLICANT: Nancy A. Hosken
; TITLE OF INVENTION: IMMUNOLOGICALLY SIGNIFICANT HERPES
; TITLE OF INVENTION: SIMPLEX VIRUS ANTIGENS AND METHODS FOR USING SAME
; FILE REFERENCE: 30967.8-US-U1
; CURRENT APPLICATION NUMBER: US/10/210,428
; CURRENT FILING DATE: 2002-07-31
; PRIOR APPLICATION NUMBER: 60/308,923
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/309,428
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Herpes simplex virus-2 (HSV-2)
US-10-210-428-1

Query Match      83.3%; Score 35; DB 2; Length 825;
Best Local Similarity 75.0%; Pred. No. 6.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      484 GAEGGPGV 491

RESULT 66
US-10-237-551-161
; Sequence 161, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 825
; TYPE: PRT
; ORGANISM: HSV-2
US-10-237-551-161

Query Match      83.3%; Score 35; DB 2; Length 826;
Best Local Similarity 75.0%; Pred. No. 6.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      485 GAEGGPGV 492

RESULT 67
US-09-894-998A-47
; Sequence 47, Application US/09894998A
; Patent No. 6537555
; GENERAL INFORMATION:
; APPLICANT: Hosken, Nancy Ann
; APPLICANT: Craig H. Day
; APPLICANT: Davin C. Dillon
; APPLICANT: McGowan, Patrick
; APPLICANT: Sleath, Paul R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538
; CURRENT APPLICATION NUMBER: US/09/894,998A
; CURRENT FILING DATE: 2001-06-28
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 826
; TYPE: PRT
; ORGANISM: HSV-2
US-09-894-998A-47

Query Match      83.3%; Score 35; DB 2; Length 826;
Best Local Similarity 75.0%; Pred. No. 6.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      485 GAEGGPGV 492

RESULT 68
US-10-237-551-47
; Sequence 47, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 826
; TYPE: PRT
; ORGANISM: HSV-2
US-10-237-551-47

Query Match      83.3%; Score 35; DB 2; Length 826;
Best Local Similarity 75.0%; Pred. No. 6.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      485 GAEGGPGV 492
```

```
RESULT 69
US-09-949-016-9530
; Sequence 9530, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9530
; LENGTH: 833
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9530

Query Match      83.3%; Score 35; DB 2; Length 833;
Best Local Similarity 75.0%; Pred. No. 6.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
    |||||
DB 630 GAEGKPGV 637

RESULT 70
US-09-949-016-6669
; Sequence 6669, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6669
; LENGTH: 966
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6669

Query Match      83.3%; Score 35; DB 2; Length 966;
Best Local Similarity 75.0%; Pred. No. 7.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAEGSPGL 8
    |||||
DB 763 GAEGKPGV 770

RESULT 71
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
```

```
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type III
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1055
; OTHER INFORMATION: /label= Modified
; OTHER INFORMATION: /note= "Ala may be Pro"
US-08-931-820-4

Query Match      83.3%; Score 35; DB 2; Length 1057;
Best Local Similarity 85.7%; Pred. No. 8.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
    |||||
DB 1025 GSEGPSG 1031

RESULT 72
US-10-153-469A-16
; Sequence 16, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUSCHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,469A
; FILING DATE: 22-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
```

;; FILING DATE: 09-OCT-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEEN, JEFFREY S
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516) 228-8484
;; TELEFAX: (516) 228-8516
;; INFORMATION FOR SEQ ID NO: 16:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1057 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-153-469A-16

Query Match 83.3%; Score 35; DB 2; Length 1057;
Best Local Similarity 85.7%; Pred. No. 8.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
||:||||
DB 591 GADGSPG 597

RESULT 73
US-10-104-889-16
; Sequence 16, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; RUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/10/104,889
;; FILING DATE: 22-Mar-2002
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/09/169,768
;; FILING DATE: 09-OCT-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEEN, JEFFREY S
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516) 228-8484
;; TELEFAX: (516) 228-8516
;; INFORMATION FOR SEQ ID NO: 16:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1057 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match 83.3%; Score 35; DB 2; Length 1057;
Best Local Similarity 85.7%; Pred. No. 8.6e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPG 7
||:||||
DB 591 GADGSPG 597

RESULT 74
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Cogoriss, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21

Query Match 83.3%; Score 35; DB 2; Length 1078;
Best Local Similarity 85.7%; Pred. No. 8.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
||:||||
DB 1026 GSEGSPPG 1032

RESULT 75
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per

```

; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
; US-09-500-811-21

```

```

Query Match      83.3%; Score 35; DB 2; Length 1078;
Best Local Similarity 85.7%; Pred. No. 8.8e+02;
Matches      6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 GAEGSPG 7
Db      1026 GSEGPSG 1032

```

```

Search completed: March 11, 2006, 12:07:26
Job time : 31.4 secs

```

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocollaboration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:06:37 ; Search time 93.6 Seconds

(without alignments)
35.712 Million cell updates/sec

Title: US-10-698-121A-1

Perfect score: 42

Sequence: 1 GAEGSPGL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications AA Main:

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	8	5	US-10-698-121A-1
2	42	100.0	12	5	US-10-698-121A-2
3	42	100.0	448	5	US-10-719-993-480
4	42	100.0	485	5	US-10-719-993-481
5	42	100.0	485	5	US-10-719-993-501
6	42	100.0	504	5	US-10-719-993-495
7	42	100.0	507	5	US-10-719-993-500
8	42	100.0	507	5	US-10-719-993-475
9	42	100.0	526	5	US-10-719-993-494
10	42	100.0	528	5	US-10-719-993-512
11	42	100.0	536	5	US-10-719-993-493
12	42	100.0	536	5	US-10-719-993-509
13	42	100.0	546	5	US-10-719-993-477
14	42	100.0	550	5	US-10-719-993-498
15	42	100.0	551	5	US-10-719-993-511
16	42	100.0	555	5	US-10-719-993-479
17	42	100.0	558	5	US-10-719-993-503
18	42	100.0	560	5	US-10-719-993-507
19	42	100.0	562	5	US-10-719-993-483
20	42	100.0	563	5	US-10-719-993-486
21	42	100.0	565	5	US-10-719-993-480
22	42	100.0	577	5	US-10-719-993-506
23	42	100.0	613	5	US-10-719-993-508
24	42	100.0	655	5	US-10-719-993-476
25	42	100.0	660	5	US-10-719-993-499
26	42	100.0	663	5	US-10-719-993-484
27	42	100.0	667	5	US-10-719-993-487

28	42	100.0	679	5	US-10-719-993-489	Sequence 489, App
29	42	100.0	691	5	US-10-719-993-491	Sequence 491, App
30	42	100.0	693	5	US-10-719-993-502	Sequence 502, App
31	42	100.0	694	5	US-10-719-993-504	Sequence 504, App
32	42	100.0	698	5	US-10-719-993-497	Sequence 497, App
33	42	100.0	701	5	US-10-719-993-492	Sequence 492, App
34	42	100.0	703	5	US-10-719-993-488	Sequence 488, App
35	42	100.0	705	5	US-10-719-993-482	Sequence 482, App
36	42	100.0	706	5	US-10-719-993-496	Sequence 496, App
37	42	100.0	708	5	US-10-719-993-485	Sequence 485, App
38	42	100.0	708	5	US-10-719-993-510	Sequence 510, App
39	42	100.0	717	5	US-10-719-993-478	Sequence 478, App
40	42	100.0	720	5	US-10-719-993-505	Sequence 505, App
41	38	90.5	219	4	US-10-104-889-44	Sequence 44, Appl
42	38	90.5	219	4	US-10-104-889-46	Sequence 46, Appl
43	38	90.5	219	5	US-10-104-793-44	Sequence 44, Appl
44	38	90.5	219	5	US-10-104-793-46	Sequence 46, Appl
45	38	90.5	498	5	US-10-901-816A-1	Sequence 1, Appl
46	38	90.5	498	5	US-10-901-816A-2	Sequence 2, Appl
47	38	90.5	660	5	US-10-901-816A-3	Sequence 3, Appl
48	38	90.5	660	5	US-10-901-816A-4	Sequence 4, Appl
49	38	90.5	698	5	US-10-732-923-1761	Sequence 1761, Ap
50	38	90.5	1014	5	US-10-901-816A-5	Sequence 5, Appl
51	38	90.5	1014	5	US-10-901-816A-6	Sequence 6, Appl
52	38	90.5	1014	5	US-10-901-816A-7	Sequence 7, Appl
53	38	90.5	1014	5	US-10-901-816A-8	Sequence 8, Appl
54	38	90.5	1014	5	US-10-901-816A-9	Sequence 9, Appl
55	38	90.5	1014	5	US-10-901-816A-10	Sequence 10, Appl
56	38	90.5	1014	5	US-10-901-816A-11	Sequence 11, Appl
57	38	90.5	1014	5	US-10-901-816A-12	Sequence 12, Appl
58	38	90.5	1014	5	US-10-901-816A-13	Sequence 13, Appl
59	38	90.5	1057	5	US-10-104-889-20	Sequence 20, Appl
60	38	90.5	1057	5	US-10-104-793-20	Sequence 20, Appl
61	38	90.5	1107	5	US-10-104-889-11	Sequence 11, Appl
62	38	90.5	1107	5	US-10-104-793-11	Sequence 11, Appl
63	38	90.5	1171	5	US-10-104-889-8	Sequence 8, Appl
64	38	90.5	1171	5	US-10-104-793-8	Sequence 8, Appl
65	38	90.5	1341	4	US-10-058-124-18	Sequence 18, Appl
66	38	90.5	1388	4	US-10-104-889-10	Sequence 10, Appl
67	38	90.5	1388	5	US-10-104-793-10	Sequence 10, Appl
68	38	90.5	1449	4	US-10-402-089-8	Sequence 8, Appl
69	38	90.5	1449	4	US-10-402-072A-8	Sequence 8, Appl
70	38	90.5	1453	4	US-10-468-091-26	Sequence 26, Appl
71	38	90.5	1461	4	US-10-468-091-25	Sequence 25, Appl
72	38	90.5	1463	4	US-10-402-089-2	Sequence 2, Appl
73	38	90.5	1463	4	US-10-402-072A-2	Sequence 2, Appl
74	38	90.5	1464	4	US-10-216-705-21	Sequence 21, Appl
75	38	90.5	1464	4	US-10-291-265-243	Sequence 243, App
76	38	90.5	1464	4	US-10-764-425-114	Sequence 114, App
77	38	90.5	1464	4	US-10-357-851-1	Sequence 1, Appl
78	38	90.5	1464	4	US-10-358-024-1	Sequence 1, Appl
79	38	90.5	1464	4	US-10-788-752-150	Sequence 150, App
80	38	90.5	1518	5	US-10-450-763-53038	Sequence 53038, A
81	37	88.1	359	4	US-10-424-599-274385	Sequence 274385, A
82	37	88.1	359	4	US-10-425-114-72009	Sequence 72009, A
83	37	88.1	359	4	US-10-425-114-72010	Sequence 72010, A
84	37	88.1	387	4	US-10-425-114-69618	Sequence 69618, A
85	37	88.1	467	4	US-10-425-114-39058	Sequence 39058, A
86	37	88.1	1027	4	US-10-360-101-221	Sequence 221, App
87	36	85.7	124	4	US-10-425-115-349811	Sequence 349811, A
88	36	85.7	340	4	US-10-156-761-9645	Sequence 9645, Ap
89	36	85.7	374	4	US-10-437-963-198005	Sequence 198005, A
90	36	85.7	416	4	US-10-029-386-33130	Sequence 33130, A
91	36	85.7	544	5	US-10-658-989A-4	Sequence 4, Appl
92	36	85.7	994	4	US-10-087-192-969	Sequence 969, App
93	36	85.7	1603	5	US-10-723-860-4225	Sequence 4225, Ap
94	36	85.7	1745	3	US-09-795-061-4	Sequence 4, Appl
95	36	85.7	1745	3	US-10-723-860-2660	Sequence 2660, Ap
96	36	85.7	1767	5	US-10-741-600-1476	Sequence 1476, Ap
97	36	85.7	1767	5	US-10-741-600-1480	Sequence 1480, Ap
98	36	85.7	1806	3	US-09-919-497-56	Sequence 56, Appl
99	36	85.7	1806	4	US-10-058-270A-122	Sequence 122, App
100	36	85.7	1806	5	US-10-741-600-1478	Sequence 1478, Ap

101	36	85.7	1806	5	US-10-741-600-1479	Sequence 1479, Ap	174	35	83.3	1464	5	US-10-723-860-2289	Sequence 2289, Ap
102	36	85.7	1806	5	US-10-287-436A-498	Sequence 198, App	176	35	83.3	1464	5	US-10-852-335A-157	Sequence 157, App
103	36	85.7	1806	5	US-10-287-436A-1192	Sequence 1192, Ap	175	35	83.3	1464	5	US-10-979-159-261	Sequence 261, App
104	36	85.7	1818	5	US-10-741-600-1475	Sequence 1475, Ap	177	35	83.3	1464	5	US-10-287-436A-491	Sequence 491, App
105	36	85.7	1818	5	US-10-741-600-1477	Sequence 1477, Ap	178	35	83.3	1464	5	US-10-287-436A-1187	Sequence 1187, App
106	35	83.3	63	6	US-11-040-130-24	Sequence 24, Appl	179	35	83.3	1466	4	US-09-918-715-226	Sequence 226, Appl
107	35	83.3	63	6	US-11-007-053-20	Sequence 20, Appl	180	35	83.3	1466	4	US-10-177-293-68	Sequence 68, Appl
108	35	83.3	97	6	US-11-097-143-19251	Sequence 19251, A	181	35	83.3	1466	4	US-10-301-822-33	Sequence 33, Appl
109	35	83.3	110	4	US-10-425-115-298923	Sequence 298923,	182	35	83.3	1466	4	US-10-257-021-72	Sequence 72, Appl
110	35	83.3	132	4	US-10-425-115-343820	Sequence 343820,	183	35	83.3	1466	4	US-10-402-089-4	Sequence 4, Appl
111	35	83.3	167	4	US-10-232-175-21	Sequence 21, Appl	184	35	83.3	1466	4	US-10-402-089-6	Sequence 6, Appl
112	35	83.3	167	6	US-11-139-377-24	Sequence 24, Appl	185	35	83.3	1466	4	US-10-402-089-12	Sequence 12, Appl
113	35	83.3	197	3	US-09-764-860-343	Sequence 343, App	186	35	83.3	1466	4	US-10-402-072A-4	Sequence 4, Appl
114	35	83.3	197	4	US-10-074-095-343	Sequence 343, App	187	35	83.3	1466	4	US-10-402-072A-6	Sequence 6, Appl
115	35	83.3	197	4	US-10-212-872-343	Sequence 343, App	188	35	83.3	1466	4	US-10-402-072A-12	Sequence 12, Appl
116	35	83.3	240	5	US-10-450-763-38730	Sequence 38730, A	189	35	83.3	1466	4	US-10-357-851-3	Sequence 3, Appl
117	35	83.3	251	4	US-10-232-175-21	Sequence 21, Appl	190	35	83.3	1466	4	US-10-358-024-3	Sequence 3, Appl
118	35	83.3	251	4	US-10-232-175-31	Sequence 31, Appl	191	35	83.3	1466	4	US-10-734-564-103	Sequence 103, App
119	35	83.3	251	6	US-11-139-377-21	Sequence 21, Appl	192	35	83.3	1466	4	US-10-474-794-226	Sequence 226, App
120	35	83.3	251	6	US-11-139-377-31	Sequence 31, Appl	193	35	83.3	1466	5	US-10-852-335A-159	Sequence 159, App
121	35	83.3	296	4	US-10-408-765A-2782	Sequence 2782, Ap	194	35	83.3	1466	5	US-10-979-159-226	Sequence 226, App
122	35	83.3	309	3	US-09-908-711-78	Sequence 78, Appl	195	35	83.3	1466	5	US-10-287-436A-451	Sequence 451, App
123	35	83.3	416	4	US-10-232-175-25	Sequence 25, Appl	196	35	83.3	1466	5	US-10-287-436A-494	Sequence 494, App
124	35	83.3	416	6	US-11-139-377-25	Sequence 25, Appl	197	35	83.3	1466	5	US-10-287-436A-1151	Sequence 1151, Ap
125	35	83.3	428	4	US-10-369-493-19121	Sequence 19121, A	198	35	83.3	1466	5	US-10-287-436A-1189	Sequence 1189, Ap
126	35	83.3	438	5	US-10-758-846-78	Sequence 78, Appl	199	35	83.3	1469	5	US-10-450-763-45550	Sequence 45550, A
127	35	83.3	459	4	US-10-408-765A-2554	Sequence 2554, Ap	200	35	83.3	1642	4	US-10-786-720-39	Sequence 39, Appl
128	35	83.3	459	5	US-10-758-846-28	Sequence 28, Appl	201	35	83.3	1670	4	US-10-786-720-38	Sequence 38, Appl
129	35	83.3	476	5	US-10-758-846-91	Sequence 91, Appl	202	35	83.3	2274	4	US-10-867-502-373	Sequence 373, App
130	35	83.3	477	5	US-10-758-846-27	Sequence 27, Appl	203	35	83.3	2274	5	US-10-805-684-140	Sequence 140, App
131	35	83.3	500	4	US-10-232-175-22	Sequence 22, Appl	204	34	81.0	39	4	US-10-767-701-60038	Sequence 60038, A
132	35	83.3	500	6	US-11-139-377-22	Sequence 22, Appl	205	34	81.0	42	4	US-10-425-115-289184	Sequence 289184, A
133	35	83.3	510	4	US-10-232-175-26	Sequence 26, Appl	206	34	81.0	63	4	US-10-437-963-110899	Sequence 110899, A
134	35	83.3	510	6	US-11-139-377-26	Sequence 26, Appl	207	34	81.0	72	4	US-10-424-599-182197	Sequence 182197, A
135	35	83.3	595	4	US-10-342-331-48	Sequence 48, Appl	208	34	81.0	78	4	US-10-425-115-263535	Sequence 263535, A
136	35	83.3	595	4	US-10-342-331-50	Sequence 50, Appl	210	34	81.0	87	4	US-10-424-599-223648	Sequence 223648, A
137	35	83.3	656	4	US-10-369-493-7782	Sequence 7782, Ap	211	34	81.0	88	5	US-10-450-763-34180	Sequence 34180, A
138	35	83.3	662	4	US-10-232-175-33	Sequence 33, Appl	212	34	81.0	92	4	US-10-437-963-133943	Sequence 133943, A
139	35	83.3	662	6	US-11-139-377-33	Sequence 33, Appl	213	34	81.0	94	4	US-10-106-698-6129	Sequence 6129, Ap
140	35	83.3	770	4	US-10-153-668-372	Sequence 372, App	214	34	81.0	97	4	US-10-424-599-170525	Sequence 170525, A
141	35	83.3	770	4	US-10-153-668-374	Sequence 374, App	215	34	81.0	99	4	US-10-029-386-32692	Sequence 32692, A
142	35	83.3	822	4	US-10-342-331-49	Sequence 49, Appl	216	34	81.0	101	4	US-10-437-963-104124	Sequence 104124, A
143	35	83.3	825	4	US-10-121-988-161	Sequence 161, App	217	34	81.0	123	4	US-10-424-599-157634	Sequence 157634, A
144	35	83.3	825	4	US-10-200-562-161	Sequence 161, App	218	34	81.0	124	4	US-10-425-115-316410	Sequence 316410, A
145	35	83.3	825	4	US-10-237-551-161	Sequence 161, App	219	34	81.0	127	4	US-10-106-698-4836	Sequence 4836, Ap
146	35	83.3	825	4	US-10-210-428-1	Sequence 1, Appl	220	34	81.0	150	3	US-09-925-298-702	Sequence 702, App
147	35	83.3	825	5	US-10-882-074-1	Sequence 1, Appl	221	34	81.0	150	4	US-10-102-806-702	Sequence 702, App
148	35	83.3	825	5	US-10-945-050-161	Sequence 161, App	222	34	81.0	163	3	US-09-992-600A-58	Sequence 58, Appl
149	35	83.3	826	4	US-09-894-998-47	Sequence 47, Appl	223	34	81.0	163	3	US-09-924-340-58	Sequence 58, Appl
150	35	83.3	826	4	US-10-121-988-47	Sequence 47, Appl	224	34	81.0	163	3	US-09-992-095B-58	Sequence 58, Appl
151	35	83.3	826	4	US-10-200-562-47	Sequence 47, Appl	225	34	81.0	163	3	US-09-999-570-58	Sequence 58, Appl
152	35	83.3	826	4	US-10-237-551-47	Sequence 47, Appl	226	34	81.0	163	4	US-10-000-986-58	Sequence 58, Appl
153	35	83.3	826	5	US-10-945-050-47	Sequence 47, App	227	34	81.0	163	4	US-10-154-678-58	Sequence 58, Appl
154	35	83.3	832	4	US-10-408-765A-1776	Sequence 1776, Ap	228	34	81.0	163	5	US-10-838-854-58	Sequence 58, Appl
155	35	83.3	966	4	US-10-153-668-256	Sequence 256, App	229	34	81.0	173	6	US-11-097-143-24567	Sequence 24567, A
156	35	83.3	966	4	US-10-153-668-258	Sequence 258, App	230	34	81.0	209	4	US-10-108-260A-4370	Sequence 4370, Ap
157	35	83.3	966	4	US-10-153-668-260	Sequence 260, App	231	34	81.0	222	4	US-10-437-963-107466	Sequence 107466, A
158	35	83.3	967	4	US-10-153-668-376	Sequence 376, App	232	34	81.0	295	4	US-10-427-631-13	Sequence 13, Appl
159	35	83.3	967	4	US-10-087-192-972	Sequence 972, App	233	34	81.0	295	4	US-10-094-097B-15	Sequence 15, Appl
160	35	83.3	1057	4	US-10-104-889-16	Sequence 16, Appl	234	34	81.0	295	4	US-10-753-267-114	Sequence 114, App
161	35	83.3	1057	5	US-10-104-793-16	Sequence 16, Appl	235	34	81.0	295	5	US-10-369-493-22783	Sequence 22783, A
162	35	83.3	1078	4	US-10-058-124-21	Sequence 21, Appl	236	34	81.0	379	4	US-10-437-963-103272	Sequence 103272, A
163	35	83.3	1169	4	US-10-104-889-6	Sequence 6, Appl	237	34	81.0	380	4	US-10-858-412-239	Sequence 239, App
164	35	83.3	1169	5	US-10-104-793-6	Sequence 6, Appl	238	34	81.0	399	5	US-09-925-302-689	Sequence 689, App
165	35	83.3	1306	4	US-10-282-122A-64405	Sequence 64405, A	239	34	81.0	403	3	US-09-925-302-689	Sequence 689, App
166	35	83.3	1464	3	US-09-918-715-261	Sequence 261, App	240	34	81.0	419	5	US-10-450-763-47660	Sequence 47660, A
167	35	83.3	1464	4	US-10-060-036-159	Sequence 159, App	241	34	81.0	433	5	US-10-858-412-240	Sequence 240, App
168	35	83.3	1464	4	US-10-171-311-36	Sequence 36, Appl	242	34	81.0	433	5	US-10-789-378-62	Sequence 62, Appl
169	35	83.3	1464	4	US-10-149-352-2	Sequence 2, Appl	243	34	81.0	483	5	US-10-758-846-88	Sequence 88, Appl
170	35	83.3	1464	4	US-10-177-293-65	Sequence 65, Appl	244	34	81.0	520	3	US-09-978-295A-614	Sequence 614, App
171	35	83.3	1464	4	US-10-301-822-28	Sequence 28, Appl	245	34	81.0	520	3	US-09-978-697-614	Sequence 614, App
172	35	83.3	1464	4	US-10-734-564-79	Sequence 79, Appl	246	34	81.0	520	3	US-09-978-192A-614	Sequence 614, App
173	35	83.3	1464	4	US-10-474-794-261	Sequence 261, App							


```
US-10-698-121A-2
; Sequence 2, Application US/10698121A
; Publication No. US20040253241A1
; GENERAL INFORMATION:
; APPLICANT: Boys Town National Research Hospital
; APPLICANT: Cosgrove, Dominic
; TITLE OF INVENTION: INDUCIBLE LIGAND FOR ALPHA1BETA1 INTEGRIN AND USSES
; FILE REFERENCE: 249,0007 0101
; CURRENT APPLICATION NUMBER: US/10/698,121A
; CURRENT FILING DATE: 2003-10-31
; PRIOR APPLICATION NUMBER: 60/423,297
; PRIOR FILING DATE: 2002-11-01
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 12
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: artificially synthesized peptide
US-10-698-121A-2
Query Match 100.0%; Score 42; DB 5; Length 12;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
Db 4 GAEGSPGL 11
RESULT 3
US-10-719-993-480
; Sequence 480, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 480
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-480
Query Match 100.0%; Score 42; DB 5; Length 448;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
Db 143 GAEGSPGL 150
RESULT 4
US-10-719-993-481
; Sequence 481, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 481
US-10-698-121A-2
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-481
Query Match 100.0%; Score 42; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
Db 124 GAEGSPGL 131
RESULT 5
US-10-719-993-501
; Sequence 501, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-501
Query Match 100.0%; Score 42; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
Db 124 GAEGSPGL 131
RESULT 6
US-10-719-993-495
; Sequence 495, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 495
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-495
Query Match 100.0%; Score 42; DB 5; Length 504;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAEGSPGL 8
Db 143 GAEGSPGL 150
RESULT 7
US-10-719-993-500
; Sequence 500, Application US/10719993
```

```
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-500

Query Match      100.0%; Score 42; DB 5; Length 507;
Best Local Similarity 100.0%; Pred. No. 90;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      146 GAEGSPGL 153

RESULT 8
US-10-719-993-475
; Sequence 475, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-475

Query Match      100.0%; Score 42; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 93;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      165 GAEGSPGL 172

RESULT 9
US-10-719-993-494
; Sequence 494, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 494
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-494

Query Match      100.0%; Score 42; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 93;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      165 GAEGSPGL 172

RESULT 10
US-10-719-993-512
; Sequence 512, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 512
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-512

Query Match      100.0%; Score 42; DB 5; Length 528;
Best Local Similarity 100.0%; Pred. No. 93;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      194 GAEGSPGL 201

RESULT 11
US-10-719-993-493
; Sequence 493, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 493
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-493

Query Match      100.0%; Score 42; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 95;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      175 GAEGSPGL 182

RESULT 12
US-10-719-993-509
; Sequence 509, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
```

```
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-509

Query Match      100.0%; Score 42; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 95;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223

RESULT 13
US-10-719-993-477
; Sequence 477, Application US/10/719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 477
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-477

Query Match      100.0%; Score 42; DB 5; Length 546;
Best Local Similarity 100.0%; Pred. No. 96;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      197 GAEGSPGL 204

RESULT 14
US-10-719-993-498
; Sequence 498, Application US/10/719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 498
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-498

Query Match      100.0%; Score 42; DB 5; Length 550;
Best Local Similarity 100.0%; Pred. No. 97;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223

; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-511

Query Match      100.0%; Score 42; DB 5; Length 551;
Best Local Similarity 100.0%; Pred. No. 97;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223

RESULT 15
US-10-719-993-511
; Sequence 511, Application US/10/719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 511
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-511

Query Match      100.0%; Score 42; DB 5; Length 551;
Best Local Similarity 100.0%; Pred. No. 97;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223

RESULT 16
US-10-719-993-479
; Sequence 479, Application US/10/719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 479
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-479

Query Match      100.0%; Score 42; DB 5; Length 555;
Best Local Similarity 100.0%; Pred. No. 98;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      194 GAEGSPGL 201

RESULT 17
US-10-719-993-503
; Sequence 503, Application US/10/719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 503
; LENGTH: 558
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-503

Query Match      100.0%; Score 42; DB 5; Length 558;
Best Local Similarity 100.0%; Pred. No. 99;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      197 GAEGSPGL 204
      |||||

RESULT 18
US-10-719-993-507
; Sequence 507, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 507
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-507

Query Match      100.0%; Score 42; DB 5; Length 560;
Best Local Similarity 100.0%; Pred. No. 99;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      199 GAEGSPGL 206
      |||||

RESULT 19
US-10-719-993-483
; Sequence 483, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 483
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-483

Query Match      100.0%; Score 42; DB 5; Length 562;
Best Local Similarity 100.0%; Pred. No. 99;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223
      |||||

RESULT 20
US-10-719-993-490
; Sequence 490, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-490

Query Match      100.0%; Score 42; DB 5; Length 563;
Best Local Similarity 100.0%; Pred. No. 99;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223
      |||||

RESULT 21
US-10-719-993-486
; Sequence 486, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 486
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-486

Query Match      100.0%; Score 42; DB 5; Length 565;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      216 GAEGSPGL 223
      |||||

RESULT 22
US-10-719-993-506
; Sequence 506, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 506
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-506

Query Match      100.0%; Score 42; DB 5; Length 577;
Best Local Similarity 100.0%; Pred. No. 1e+02;
```



```
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 216 GAEGSPGL 223

RESULT 23
US-10-719-993-508
; Sequence 508, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 508
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-508

Query Match 100.0%; Score 42; DB 5; Length 613;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 308 GAEGSPGL 315

RESULT 24
US-10-719-993-476
; Sequence 476, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 476
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-476

Query Match 100.0%; Score 42; DB 5; Length 655;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 294 GAEGSPGL 301

RESULT 25
US-10-719-993-499
; Sequence 499, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
```

```
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 499
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-499

Query Match 100.0%; Score 42; DB 5; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 299 GAEGSPGL 306

RESULT 26
US-10-719-993-484
; Sequence 484, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 484
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-484

Query Match 100.0%; Score 42; DB 5; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 302 GAEGSPGL 309

RESULT 27
US-10-719-993-487
; Sequence 487, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 487
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-487

Query Match 100.0%; Score 42; DB 5; Length 667;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
    |||||
Db 306 GAEGSPGL 313
```

```
RESULT 28
US-10-719-993-489
; Sequence 489, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489
; LENGTH: 679
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-489

Query Match      100.0%; Score 42; DB 5; Length 679;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      318 GAEGSPGL 325

RESULT 29
US-10-719-993-491
; Sequence 491, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 491
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-491

Query Match      100.0%; Score 42; DB 5; Length 691;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      330 GAEGSPGL 337

RESULT 30
US-10-719-993-502
; Sequence 502, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 502
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-502

Query Match      100.0%; Score 42; DB 5; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 31
US-10-719-993-504
; Sequence 504, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 504
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-504

Query Match      100.0%; Score 42; DB 5; Length 694;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 32
US-10-719-993-497
; Sequence 497, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 497
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-497

Query Match      100.0%; Score 42; DB 5; Length 698;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      337 GAEGSPGL 344

RESULT 33
US-10-719-993-492
; Sequence 492, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
```

```
; ORGANISM: Homo sapiens
US-10-719-993-502

Query Match      100.0%; Score 42; DB 5; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 31
US-10-719-993-504
; Sequence 504, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 504
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-504

Query Match      100.0%; Score 42; DB 5; Length 694;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 32
US-10-719-993-497
; Sequence 497, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 497
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-497

Query Match      100.0%; Score 42; DB 5; Length 698;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      337 GAEGSPGL 344

RESULT 33
US-10-719-993-492
; Sequence 492, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
```

```

; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 492
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-492

Query Match      100.0%; Score 42; DB 5; Length 701;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      340 GAEGSPGL 347

RESULT 34
US-10-719-993-488
; Sequence 488, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 703
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-488

Query Match      100.0%; Score 42; DB 5; Length 703;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      342 GAEGSPGL 349

RESULT 35
US-10-719-993-482
; Sequence 482, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 482
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-482

Query Match      100.0%; Score 42; DB 5; Length 705;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      342 GAEGSPGL 349

RESULT 36
US-10-719-993-496
; Sequence 496, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 496
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-496

Query Match      100.0%; Score 42; DB 5; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 37
US-10-719-993-485
; Sequence 485, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 485
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-485

Query Match      100.0%; Score 42; DB 5; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      359 GAEGSPGL 366

RESULT 38
US-10-719-993-510
; Sequence 510, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
```

; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 510
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-510

Query Match 100.0%; Score 42; DB 5; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 347 GAEGSPGL 354

RESULT 39

US-10-719-993-478
; Sequence 478, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 478
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-478

Query Match 100.0%; Score 42; DB 5; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 356 GAEGSPGL 363

RESULT 40

US-10-719-993-505
; Sequence 505, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 505
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-505

Query Match 100.0%; Score 42; DB 5; Length 720;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 359 GAEGSPGL 366

RESULT 41

US-10-104-889-44
; Sequence 44, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; BUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-104-889-44

Query Match 90.5%; Score 38; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26

RESULT 42

US-10-104-889-46
; Sequence 46, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; BUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.

ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-104-889-46

Query Match 90.5%; Score 38; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26

RESULT 43
US-10-104-793-44
; Sequence 44, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; BUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,793
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-104-793-46

Query Match 90.5%; Score 38; DB 5; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26

INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 44:
US-10-104-793-44

Query Match 90.5%; Score 38; DB 5; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26

RESULT 44
US-10-104-793-46
; Sequence 46, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; BUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,793
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-10-104-793-46

Query Match 90.5%; Score 38; DB 5; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 20 GAEGSPG 26

```
RESULT 45
US-10-901-816A-1
; Sequence 1, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-1

Query Match          90.5%; Score 38; DB 5; Length 498;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493

RESULT 46
US-10-901-816A-2
; Sequence 2, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-2

Query Match          90.5%; Score 38; DB 5; Length 498;
Best Local Similarity 100.0%; Pred. No. 4.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493

RESULT 47
US-10-901-816A-3
; Sequence 3, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 1761
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Bactrocera tryoni
US-10-901-816A-3

Query Match          90.5%; Score 38; DB 5; Length 660;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493

RESULT 48
US-10-901-816A-4
; Sequence 4, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-4

Query Match          90.5%; Score 38; DB 5; Length 660;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
Db 487 GAEGSPG 493

RESULT 49
US-10-732-923-1761
; Sequence 1761, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154
; PRIOR FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 24149
; SEQ ID NO 1761
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Bactrocera tryoni
US-10-732-923-1761
```

US-10-732-923-1761

Query Match 90.5%; Score 38; DB 5; Length 698;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||
Db 40 GAEGSPG 46

RESULT 50

US-10-901-816A-5
; Sequence 5, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-5

Query Match 90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||
Db 841 GAEGSPG 847

RESULT 51

US-10-901-816A-6
; Sequence 6, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-6

Query Match 90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||
Db 841 GAEGSPG 847

RESULT 52

US-10-901-816A-7
; Sequence 7, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-7

Query Match 90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||
Db 841 GAEGSPG 847

RESULT 53

US-10-901-816A-8
; Sequence 8, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901,816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-8

Query Match 90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
|||||
Db 841 GAEGSPG 847

RESULT 54

US-10-901-816A-9
; Sequence 9, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.


```
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-9

Query Match          90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      841 GAEGSPG 847

RESULT 55
US-10-901-816A-10
; Sequence 10, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-10

Query Match          90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      841 GAEGSPG 847

RESULT 56
US-10-901-816A-11
; Sequence 11, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-13

Query Match          90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      841 GAEGSPG 847

RESULT 57
US-10-901-816A-12
; Sequence 12, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-12

Query Match          90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      841 GAEGSPG 847

RESULT 58
US-10-901-816A-13
; Sequence 13, Application US/10901816A
; Publication No. US20050058703A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Robert C.
; APPLICANT: Olsen, David R.
; APPLICANT: James, Polarek W.
; APPLICANT: Williams, Kim E.
; TITLE OF INVENTION: Gelatin Capsules
; FILE REFERENCE: FP0404 US
; CURRENT APPLICATION NUMBER: US/10/901.816A
; CURRENT FILING DATE: 2004-07-29
; PRIOR APPLICATION NUMBER: US 60/492,085
; PRIOR FILING DATE: 2003-08-01
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1014
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-901-816A-13

Query Match          90.5%; Score 38; DB 5; Length 1014;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 841 GAEGSPG 847

RESULT 59

US-10-104-889-20

; Sequence 20, Application US/10104889

; Publication No. US20040086961A1

; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.

BUECHTER, DOUGLAS

BROKAW, JANE

ZHANG, GUANGHUI

PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889

FILING DATE: 22-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768

FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1057 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-10-104-889-20

Query Match 90.5%; Score 38; DB 4; Length 1057;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 858 GAEGSPG 864

RESULT 60

US-10-104-793-20

; Sequence 20, Application US/10104793

; Publication No. US20050196830A1

; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.

BUECHTER, DOUGLAS

BROKAW, JANE

ZHANG, GUANGHUI

PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889

FILING DATE: 22-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768

FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1057 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-10-104-793-20

Query Match 90.5%; Score 38; DB 5; Length 1057;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 858 GAEGSPG 864

RESULT 61

US-10-104-889-11

; Sequence 11, Application US/10104889

; Publication No. US20040086961A1

; GENERAL INFORMATION:

APPLICANT: GRUSKIN, ELLIOT A.

BUECHTER, DOUGLAS

BROKAW, JANE

ZHANG, GUANGHUI

PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: DILWORTH & BARRESE

STREET: 333 EARLE OVINGTON BOULEVARD

CITY: UNIONDALE

STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/104,889

FILING DATE: 22-Mar-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768

FILING DATE: 09-OCT-1998

ATTORNEY/AGENT INFORMATION:

NAME: STEEN, JEFFREY S

TELECOMMUNICATION INFORMATION:

TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1057 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-10-104-793-20

Query Match 90.5%; Score 38; DB 5; Length 1057;

Best Local Similarity 100.0%; Pred. No. 8.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 858 GAEGSPG 864

; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-889-11

Query Match 90.5%; Score 38; DB 4; Length 1107;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864
|||||

RESULT 62
US-10-104-793-11
; Sequence 11, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,793
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-793-11

Query Match 90.5%; Score 38; DB 5; Length 1107;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864
|||||

RESULT 62
US-10-104-793-11
; Sequence 11, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,793
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1107 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-793-11

Query Match 90.5%; Score 38; DB 5; Length 1107;
Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864
|||||

RESULT 63
US-10-104-889-8
; Sequence 8, Application US/10104889
; Publication No. US20040086961A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-889-8

Query Match 90.5%; Score 38; DB 4; Length 1171;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864
|||||

RESULT 64
US-10-104-793-8
; Sequence 8, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID

;; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
;; NUMBER OF SEQUENCES: 50
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: DILWORTH & BARRESE
;; STREET: 333 EARLE OVINGTON BOULEVARD
;; CITY: UNIONDALE
;; STATE: NY
;; COUNTRY: U.S.A.
;; ZIP: 11553
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA: US/10/104,793
;; APPLICATION NUMBER: US/10/104,793
;; FILING DATE: 22-Mar-2002
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/09/169,768
;; FILING DATE: 09-OCT-1998
;; ATTORNEY/AGENT INFORMATION:
;; NAME: STEEN, JEFFREY S
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (516) 228-8484
;; TELEFAX: (516) 228-8516
;;
;; INFORMATION FOR SEQ ID NO: 8:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1171 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-104-793-8

Query Match 90.5%; Score 38; DB 5; Length 1171;
Best Local Similarity 100.0%; Pred. No. 9.3e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 1 GAEGSPG 7
Db 858 GAEGSPG 864

RESULT 65

US-10-058-124-18
; Sequence 18, Application US/10058124
; Publication No. US20030119058A1

GENERAL INFORMATION:

APPLICANT: Qvist, Per
Bonde, Martin

TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/570,573
;; FILING DATE: 2002-MAY-12
;; APPLICATION NUMBER: 08/187,319
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gogoris, Adda C
;; REGISTRATION NUMBER: 29,714
;; REFERENCE/DOCKET NUMBER: 4305/08701
;; TELEPHONE: 212-527-7700
;; TELEFAX: 212-753-6237
;; TELEX: 236687
;;
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1341 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; IMMEDIATE SOURCE:
;; CLONE: COLLAGEN ALPHA 1 (I)
;; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-058-124-18

Query Match 90.5%; Score 38; DB 4; Length 1341;
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

Qy 1 GAEGSPG 7
Db 895 GAEGSPG 901

RESULT 66

US-10-104-889-10
; Sequence 10, Application US/10104889
; Publication No. US20040086961A1

GENERAL INFORMATION:

APPLICANT: BRUSKIN, ELLIOT A.
BROKAW, JANE
BUECHTER, DOUGLAS
ZHANG, GUANGHUI
PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY

COUNTRY: U.S.A.

ZIP: 11553

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:

```
;
; LENGTH: 1388 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-889-10

Query Match          90.5%; Score 38; DB 4; Length 1388;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||||
Db 858 GAEGSPG 864

RESULT 67
US-10-104-793-10
; Sequence 10, Application US/10104793
; Publication No. US20050196830A1
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1388 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-10-104-793-10

Query Match          90.5%; Score 38; DB 5; Length 1388;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||||
Db 858 GAEGSPG 864

RESULT 68
```

```
US-10-402-089-8
; Sequence 8, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-8

Query Match          90.5%; Score 38; DB 4; Length 1449;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||||
Db 1004 GAEGSPG 1010

RESULT 69
US-10-402-072A-8
; Sequence 8, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 8
; LENGTH: 1449
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-8

Query Match          90.5%; Score 38; DB 4; Length 1449;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||||
Db 1004 GAEGSPG 1010

RESULT 70
US-10-468-091-26
; Sequence 26, Application US/10468091
; Publication No. US20040157329A1
; GENERAL INFORMATION:
; APPLICANT: ADP Pharmaceutical Pty Limited
; APPLICANT: The University of Sydney
; TITLE OF INVENTION: Matrix gene expression in chondrogenesis
; FILE REFERENCE: 500311
; CURRENT APPLICATION NUMBER: US/10/468,091
```

; CURRENT FILING DATE: 2003-08-13
; PRIOR APPLICATION NUMBER: AU PR3116
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 1453
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-468-091-26

Query Match 90.5%; Score 38; DB 4; Length 1453;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
| | | | |
DB 1008 GAEGSPG 1014

RESULT 71

US-10-468-091-25
; Sequence 25, Application US/10468091
; Publication No. US20040157329A1
; GENERAL INFORMATION:
; APPLICANT: ADP Pharmaceutical Pty Limited
; APPLICANT: The University of Sydney
; TITLE OF INVENTION: Matrix gene expression in chondrogenesis
; FILE REFERENCE: 500311
; CURRENT APPLICATION NUMBER: US/10/468,091
; CURRENT FILING DATE: 2003-08-13
; PRIOR APPLICATION NUMBER: AU PR3116
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-468-091-25

Query Match 90.5%; Score 38; DB 4; Length 1461;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
| | | | |
DB 1016 GAEGSPG 1022

RESULT 72

US-10-402-089-2
; Sequence 2, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcus P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Sealey, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-089-2

Query Match 90.5%; Score 38; DB 4; Length 1463;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
| | | | |
DB 1018 GAEGSPG 1024

RESULT 73

US-10-402-072A-2
; Sequence 2, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcus P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Sealey, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1463
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-072A-2

Query Match 90.5%; Score 38; DB 4; Length 1463;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
| | | | |
DB 1018 GAEGSPG 1024

RESULT 74

US-10-216-705-21
; Sequence 21, Application US/10216705
; Publication No. US20030096973A1
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, M
; FILE REFERENCE: 1149-3 DIV
; CURRENT APPLICATION NUMBER: US/10/216,705
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 09/331,347
; PRIOR FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-216-705-21

Query Match 90.5%; Score 38; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAEGSPG 7
| | | | |
DB 1019 GAEGSPG 1025

RESULT 75

US-10-291-265-243

```

; Sequence 243, Application US/10291265
; Publication No. US20030232054A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang et al
; TITLE OF INVENTION: No. US20030232054A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-017 (785)
; CURRENT APPLICATION NUMBER: US/10/291,265
; CURRENT FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-291-265-243

Query Match          90.5%; Score 38; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
Db      1019 GAEGSPG 1025

```

Search completed: March 11, 2006, 12:11:28
Job time : 101.6 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:07:47 ; Search time 11.2 Seconds
(without alignments)
19.882 Million cell updates/sec

Title: US-10-698-121A-1

Perfect score: 42

Sequence: 1 GAEGSPGL 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 161667 seqs, 27834885 residues

Total number of hits satisfying chosen parameters: 161667

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	90.5	317	7 US-11-024-959-335	Sequence 335, App
2	38	90.5	1464	7 US-11-000-463-243	Sequence 243, App
3	36	85.7	749	7 US-11-052-554A-148	Sequence 148, App
4	36	85.7	1081	7 US-11-051-720-1372	Sequence 1372, App
5	36	85.7	1519	6 US-11-182-016-5	Sequence 5, Appli
6	36	85.7	1767	6 US-10-995-561-911	Sequence 911, App
7	36	85.7	1767	6 US-10-995-561-914	Sequence 914, App
8	36	85.7	1806	6 US-10-995-561-912	Sequence 912, App
9	36	85.7	1806	6 US-10-995-561-915	Sequence 915, App
10	36	85.7	1806	7 US-11-051-720-1446	Sequence 1446, App
11	36	85.7	1806	7 US-11-051-720-1447	Sequence 1447, App
12	36	85.7	1818	6 US-10-995-561-910	Sequence 910, App
13	36	85.7	1818	6 US-10-995-561-913	Sequence 913, App
14	35	83.3	150	7 US-11-096-568A-11972	Sequence 11972, A
15	35	83.3	260	6 US-10-485-517-354	Sequence 354, App
16	35	83.3	261	6 US-10-485-517-150	Sequence 150, App
17	35	83.3	1166	6 US-10-821-234-964	Sequence 964, App
18	35	83.3	1306	7 US-11-052-554A-139	Sequence 139, App
19	35	83.3	1464	6 US-10-501-035-331	Sequence 331, App
20	35	83.3	1464	7 US-11-186-284-28	Sequence 28, Appli
21	35	83.3	1464	7 US-11-021-603-2	Sequence 2, Appli
22	35	83.3	1466	7 US-11-186-284-33	Sequence 33, Appl
23	35	83.3	1467	6 US-10-821-234-1096	Sequence 1096, App
24	34	81.0	295	7 US-11-193-205-1	Sequence 1, Appli
25	34	81.0	520	6 US-10-995-561-532	Sequence 532, App

26	34	81.0	571	7 US-11-072-512-3814	Sequence 3814, Ap
27	34	81.0	618	6 US-10-821-234-1481	Sequence 1481, Ap
28	34	81.0	822	6 US-10-330-773-700	Sequence 700, App
29	33	78.6	288	7 US-11-135-855-30	Sequence 30, Appl
30	33	78.6	289	7 US-11-258-647-15	Sequence 15, Appl
31	33	78.6	303	7 US-11-135-855-31	Sequence 31, Appl
32	33	78.6	303	7 US-11-258-647-2	Sequence 2, Appli
33	33	78.6	756	7 US-11-072-512-3505	Sequence 2505, Ap
34	33	78.6	828	6 US-10-995-561-983	Sequence 983, App
35	33	78.6	918	6 US-10-995-561-981	Sequence 981, App
36	33	78.6	1019	6 US-10-995-561-982	Sequence 982, App
37	32	76.2	117	7 US-11-096-070-34	Sequence 34, Appl
38	32	76.2	164	7 US-11-087-227-18	Sequence 18, Appl
39	32	76.2	164	7 US-11-087-227-20	Sequence 20, Appl
40	32	76.2	179	7 US-11-197-133A-62	Sequence 62, Appl
41	32	76.2	288	6 US-10-131-826A-316	Sequence 316, App
42	32	76.2	288	6 US-10-973-115B-316	Sequence 316, App
43	32	76.2	295	6 US-10-995-561-999	Sequence 999, App
44	32	76.2	299	7 US-11-177-506-37	Sequence 37, Appl
45	32	76.2	350	7 US-11-087-099-12163	Sequence 12163, A
46	32	76.2	406	7 US-11-072-512-3291	Sequence 3291, Ap
47	32	76.2	438	7 US-11-072-512-3166	Sequence 3166, Ap
48	32	76.2	531	7 US-11-096-070-4	Sequence 4, Appli
49	32	76.2	549	7 US-11-096-070-6	Sequence 6, Appli
50	32	76.2	549	7 US-11-096-070-8	Sequence 8, Appli
51	32	76.2	576	7 US-11-052-554A-164	Sequence 164, App
52	32	76.2	615	7 US-11-052-554A-152	Sequence 152, App
53	32	76.2	737	6 US-10-501-035-254	Sequence 254, App
54	32	76.2	801	7 US-11-052-554A-166	Sequence 166, App
55	32	76.2	837	7 US-11-052-554A-159	Sequence 159, App
56	32	76.2	861	7 US-11-186-641A-2	Sequence 2, Appli
57	32	76.2	914	6 US-11-052-554A-160	Sequence 160, App
58	32	76.2	915	6 US-10-821-234-1514	Sequence 1514, Ap
59	32	76.2	915	6 US-10-995-561-1003	Sequence 1003, Ap
60	32	76.2	915	6 US-10-995-561-1000	Sequence 1000, Ap
61	32	76.2	940	6 US-10-995-561-1004	Sequence 1004, Ap
62	32	76.2	969	6 US-10-995-561-1001	Sequence 1001, Ap
63	32	76.2	971	6 US-10-995-561-998	Sequence 998, App
64	32	76.2	994	6 US-10-995-561-997	Sequence 997, App
65	32	76.2	1366	6 US-10-821-234-1431	Sequence 1431, Ap
66	32	76.2	1366	7 US-11-186-284-31	Sequence 31, Appl
67	32	76.2	1736	7 US-11-124-368A-329	Sequence 329, App
68	32	76.2	1783	7 US-11-126-313-38	Sequence 38, Appl
69	31	73.8	76	7 US-11-096-568A-12579	Sequence 8268, Ap
70	31	73.8	219	7 US-11-087-099-8268	Sequence 8268, Ap
71	31	73.8	285	7 US-11-072-512-3028	Sequence 3028, Ap
72	31	73.8	388	6 US-10-858-730-83	Sequence 83, Appl
73	31	73.8	414	7 US-11-115-868-2	Sequence 2, Appli
74	31	73.8	414	7 US-11-165-305-2	Sequence 2, Appli
75	31	73.8	542	7 US-11-074-176-30	Sequence 30, Appl
76	31	73.8	711	6 US-10-517-939-4	Sequence 4, Appli
77	31	73.8	778	7 US-11-052-554A-144	Sequence 144, App
78	31	73.8	1874	6 US-10-821-234-1182	Sequence 1182, Ap
79	31	73.8	1971	7 US-11-179-624-1	Sequence 1, Appli
80	31	73.8	3132	7 US-11-087-099-1245	Sequence 1245, Ap
81	31	73.8	129	6 US-10-967-527A-13	Sequence 13, Appl
82	30	71.4	156	7 US-11-087-099-5494	Sequence 5494, Ap
83	30	71.4	177	6 US-10-999-866-61	Sequence 61, Appl
84	30	71.4	189	7 US-11-096-568A-10728	Sequence 10728, A
85	30	71.4	205	6 US-10-995-561-1028	Sequence 1028, Ap
86	30	71.4	227	6 US-10-995-561-1029	Sequence 1029, Ap
87	30	71.4	243	6 US-10-063-703-122	Sequence 122, App
88	30	71.4	243	7 US-11-102-240-122	Sequence 122, App
89	30	71.4	243	7 US-11-080-991-14	Sequence 14, Appl
90	30	71.4	255	7 US-11-115-086-11	Sequence 11, Appl
91	30	71.4	270	7 US-11-096-568A-25048	Sequence 25048, A
92	30	71.4	278	7 US-11-186-284-77	Sequence 77, Appl
93	30	71.4	280	5 US-09-978-360A-809	Sequence 809, App
94	30	71.4	280	6 US-10-821-234-1300	Sequence 1300, Ap
95	30	71.4	287	7 US-11-096-568A-10726	Sequence 10726, A

99	30	71.4	307	6	US-10-858-730-84	Sequence 84, Appl	172	29	69.0	303	6	US-10-922-232B-58	Sequence 58, Appl
100	30	71.4	314	7	US-11-087-099-2664	Sequence 2664, Ap	173	29	69.0	309	7	US-11-096-568A-18960	Sequence 18960, A
101	30	71.4	316	7	US-11-087-099-778	Sequence 778, App	174	29	69.0	314	7	US-11-096-568A-26059	Sequence 26059, A
102	30	71.4	316	7	US-11-096-568A-24774	Sequence 24774, A	175	29	69.0	318	6	US-10-802-796-727	Sequence 727, App
103	30	71.4	354	7	US-11-096-568A-23336	Sequence 23336, A	176	29	69.0	318	6	US-11-060-029-21	Sequence 21, Appl
104	30	71.4	376	7	US-11-219-282-28	Sequence 28, Appl	177	29	69.0	326	6	US-10-055-877-289	Sequence 289, App
105	30	71.4	377	7	US-11-124-368A-209	Sequence 209, App	178	29	69.0	326	6	US-10-055-877-297	Sequence 297, App
106	30	71.4	377	7	US-11-124-368A-210	Sequence 210, App	179	29	69.0	328	7	US-11-096-568A-25744	Sequence 25744, A
107	30	71.4	407	7	US-11-072-512-3895	Sequence 3895, Ap	180	29	69.0	332	7	US-11-096-568A-8160	Sequence 8160, Ap
108	30	71.4	522	7	US-11-087-099-6999	Sequence 6999, Ap	181	29	69.0	334	6	US-10-802-796-728	Sequence 728, App
109	30	71.4	551	7	US-11-219-282-27	Sequence 27, Appl	182	29	69.0	337	6	US-10-063-703-74	Sequence 74, Appl
110	30	71.4	638	7	US-11-100-640-14	Sequence 14, Appl	183	29	69.0	337	7	US-11-102-240-74	Sequence 74, Appl
111	30	71.4	677	6	US-10-982-545-12	Sequence 12, Appl	184	29	69.0	337	7	US-11-087-099-5945	Sequence 5945, Ap
112	30	71.4	729	7	US-11-051-720-1373	Sequence 1373, Ap	185	29	69.0	344	7	US-11-060-029-15	Sequence 15, Appl
113	30	71.4	744	7	US-11-186-284-37	Sequence 37, Appl	186	29	69.0	345	7	US-11-096-568A-10730	Sequence 10730, A
114	30	71.4	744	7	US-11-186-284-39	Sequence 39, Appl	187	29	69.0	346	7	US-11-060-029-19	Sequence 19, Appl
115	30	71.4	783	7	US-11-052-554A-157	Sequence 157, App	188	29	69.0	351	7	US-11-147-849-1	Sequence 1, Appli
116	30	71.4	834	6	US-10-131-826A-148	Sequence 148, App	189	29	69.0	351	7	US-11-147-711-1	Sequence 1, Appli
117	30	71.4	834	6	US-10-973-115B-148	Sequence 148, App	190	29	69.0	373	6	US-10-793-626-890	Sequence 890, App
118	30	71.4	834	7	US-11-072-512-2323	Sequence 2323, Ap	191	29	69.0	378	7	US-11-096-568A-26057	Sequence 26057, A
119	30	71.4	841	6	US-10-624-932-6	Sequence 6, Appli	192	29	69.0	394	7	US-11-072-512-3560	Sequence 3560, Ap
120	30	71.4	841	6	US-10-624-932-8	Sequence 8, Appli	193	29	69.0	397	7	US-11-096-568A-20883	Sequence 20883, A
121	30	71.4	897	7	US-11-124-367A-449	Sequence 449, App	194	29	69.0	398	7	US-11-190-188-5	Sequence 5, Appli
122	30	71.4	897	7	US-11-124-367A-451	Sequence 451, App	195	29	69.0	399	7	US-11-096-568A-15884	Sequence 15884, A
123	30	71.4	903	7	US-11-124-367A-450	Sequence 450, App	196	29	69.0	400	6	US-10-689-742-74	Sequence 74, Appl
124	30	71.4	922	7	US-11-115-086-9	Sequence 9, Appli	197	29	69.0	405	7	US-11-087-099-3071	Sequence 3071, Ap
125	30	71.4	974	7	US-11-115-086-7	Sequence 7, Appli	198	29	69.0	406	6	US-10-131-826A-82	Sequence 82, Appl
126	30	71.4	1028	7	US-11-169-041-180	Sequence 180, App	199	29	69.0	406	6	US-10-973-115B-82	Sequence 82, Appl
127	30	71.4	1496	7	US-11-186-284-35	Sequence 35, Appl	200	29	69.0	427	7	US-11-096-070-10	Sequence 10, Appl
128	30	71.4	1516	6	US-10-220-824-8	Sequence 8, Appli	201	29	69.0	436	7	US-11-087-099-383	Sequence 383, App
129	30	71.4	1532	6	US-10-821-234-914	Sequence 914, App	202	29	69.0	441	7	US-11-100-640-30	Sequence 30, Appl
130	30	71.4	3063	7	US-11-186-284-26	Sequence 26, Appl	203	29	69.0	442	7	US-11-096-568A-15883	Sequence 15883, A
131	30	71.4	8746	7	US-11-088-686-10232	Sequence 10232, A	204	29	69.0	449	7	US-11-096-568A-15882	Sequence 15882, A
132	29	69.0	11	6	US-11-103-708-4	Sequence 4, Appli	205	29	69.0	467	7	US-11-087-099-9910	Sequence 9910, Ap
133	29	69.0	15	6	US-10-939-890-491	Sequence 491, App	206	29	69.0	485	7	US-11-096-568A-11820	Sequence 11820, A
134	29	69.0	40	6	US-10-979-871-14	Sequence 14, Appl	207	29	69.0	507	7	US-11-087-099-3383	Sequence 3383, Ap
135	29	69.0	40	6	US-10-979-871-18	Sequence 18, Appl	208	29	69.0	508	7	US-11-174-398-16	Sequence 16, Appl
136	29	69.0	79	7	US-11-245-689-9	Sequence 9, Appli	209	29	69.0	538	7	US-11-124-368A-311	Sequence 311, App
137	29	69.0	85	6	US-10-986-501-261	Sequence 261, App	210	29	69.0	538	7	US-11-124-368A-312	Sequence 312, App
138	29	69.0	102	7	US-11-245-689-8	Sequence 8, Appli	211	29	69.0	538	7	US-11-124-368A-313	Sequence 313, App
139	29	69.0	114	7	US-11-245-689-7	Sequence 7, Appli	212	29	69.0	538	7	US-11-132-947-6	Sequence 6, Appli
140	29	69.0	114	7	US-11-245-689-21	Sequence 21, Appl	213	29	69.0	538	7	US-11-197-488-2	Sequence 2, Appli
141	29	69.0	117	7	US-11-087-099-5641	Sequence 5641, Ap	214	29	69.0	551	7	US-11-096-070-12	Sequence 12, Appl
142	29	69.0	118	7	US-11-096-568A-15267	Sequence 15267, A	215	29	69.0	629	6	US-10-821-234-1528	Sequence 1528, Ap
143	29	69.0	120	7	US-11-086-070-35	Sequence 35, Appl	216	29	69.0	652	6	US-10-821-234-1015	Sequence 1016, Ap
144	29	69.0	135	7	US-11-245-689-11	Sequence 11, Appl	217	29	69.0	682	7	US-11-072-175-154	Sequence 154, App
145	29	69.0	138	6	US-10-821-234-1172	Sequence 1172, Ap	218	29	69.0	695	6	US-10-363-924-2	Sequence 2, Appli
146	29	69.0	138	6	US-11-245-689-16	Sequence 16, Appl	219	29	69.0	720	7	US-11-113-424-28	Sequence 28, Appl
147	29	69.0	162	6	US-10-467-657-9046	Sequence 9046, Ap	220	29	69.0	738	7	US-11-051-720-1374	Sequence 1374, Ap
148	29	69.0	162	7	US-11-024-959-441	Sequence 441, App	221	29	69.0	741	7	US-11-052-554A-161	Sequence 161, App
149	29	69.0	165	7	US-11-245-689-20	Sequence 20, Appl	222	29	69.0	747	7	US-11-113-424-26	Sequence 26, Appl
150	29	69.0	171	7	US-11-072-512-3688	Sequence 3688, Ap	223	29	69.0	767	7	US-11-052-554A-154	Sequence 154, App
151	29	69.0	192	7	US-11-096-568A-17201	Sequence 17201, A	224	29	69.0	813	7	US-11-087-099-4670	Sequence 4670, Ap
152	29	69.0	194	7	US-11-072-512-3133	Sequence 3133, Ap	225	29	69.0	820	7	US-11-165-819-1	Sequence 1, Appli
153	29	69.0	224	7	US-11-096-568A-17200	Sequence 17200, A	226	29	69.0	824	6	US-10-821-234-1008	Sequence 1008, Ap
154	29	69.0	243	6	US-10-131-826A-362	Sequence 362, App	227	29	69.0	853	7	US-11-052-554A-149	Sequence 149, App
155	29	69.0	243	6	US-10-973-115B-362	Sequence 362, App	228	29	69.0	1068	6	US-10-453-372-1084	Sequence 1084, Ap
156	29	69.0	243	6	US-11-072-512-3694	Sequence 3694, Ap	229	29	69.0	1068	6	US-10-453-372-1090	Sequence 1090, Ap
157	29	69.0	244	6	US-10-514-040-4	Sequence 4, Appli	230	29	69.0	1077	6	US-10-453-372-1086	Sequence 1086, Ap
158	29	69.0	244	6	US-10-296-865-6	Sequence 6, Appli	231	29	69.0	1093	6	US-10-453-372-1088	Sequence 1088, Ap
159	29	69.0	244	7	US-11-157-947-1	Sequence 1, Appli	232	29	69.0	1124	7	US-11-195-197-9	Sequence 9, Appli
160	29	69.0	244	7	US-11-256-802-3	Sequence 3, Appli	233	29	69.0	1124	7	US-11-049-536-162	Sequence 162, App
161	29	69.0	244	7	US-11-258-647-4	Sequence 4, Appli	234	29	69.0	1124	7	US-11-114-379-1	Sequence 1, Appli
162	29	69.0	245	7	US-11-256-802-4	Sequence 4, Appli	235	29	69.0	1329	7	US-11-052-554A-136	Sequence 136, App
163	29	69.0	252	7	US-11-096-568A-25298	Sequence 25298, A	236	29	69.0	1329	7	US-11-087-099-882	Sequence 882, App
164	29	69.0	253	6	US-10-821-234-1438	Sequence 1438, Ap	237	29	69.0	1381	7	US-11-052-554A-138	Sequence 138, App
165	29	69.0	263	7	US-11-087-099-6263	Sequence 6263, Ap	238	29	69.0	1408	7	US-11-087-099-8482	Sequence 8482, Ap
166	29	69.0	271	7	US-11-096-568A-17199	Sequence 17199, A	239	29	69.0	1432	6	US-10-510-386-218	Sequence 218, App
167	29	69.0	282	7	US-11-185-739-11	Sequence 11, Appl	240	29	69.0	1823	6	US-10-995-561-988	Sequence 988, App
168	29	69.0	282	7	US-11-096-568A-14838	Sequence 14838, A	241	29	69.0	1901	7	US-11-052-554A-135	Sequence 135, App
169	29	69.0	282	7	US-11-096-568A-25745	Sequence 25745, A	242	29	69.0	1970	6	US-10-821-234-1641	Sequence 1641, Ap
170	29	69.0	285	7	US-11-258-647-5	Sequence 5, Appli	243	29	69.0	2102	6	US-10-995-561-990	Sequence 990, App
171	29	69.0	299	7	US-11-096-568A-10731	Sequence 10731, A	244	29	69.0	2108	6	US-10-995-561-989	Sequence 989, App

245	29	69.0	2157	6	US-10-995-561-991	Sequence 991, App	318	28	66.7	689	7	US-11-087-099-911	Sequence 911, App
246	29	69.0	2923	7	US-11-200-822-3	Sequence 3, Appli	319	28	66.7	691	7	US-11-087-099-4572	Sequence 4572, Ap
247	29	69.0	4060	6	US-10-922-232B-55	Sequence 55, Appl	320	28	66.7	694	7	US-11-052-554A-158	Sequence 158, App
248	29	69.0	6738	6	US-10-922-232B-56	Sequence 56, Appl	321	28	66.7	709	6	US-10-453-372-182	Sequence 182, App
249	29	69.0	7465	7	US-11-087-099-7521	Sequence 7521, Ap	322	28	66.7	709	6	US-10-453-372-184	Sequence 184, App
250	29	69.0	16990	7	US-11-175-689-7	Sequence 2, Appli	323	28	66.7	709	6	US-10-453-372-186	Sequence 186, App
251	28	66.7	11	7	US-11-105-708-2	Sequence 2, Appli	324	28	66.7	729	6	US-11-096-568A-19858	Sequence 19858, A
252	28	66.7	109	7	US-11-064-174-57	Sequence 57, Appl	325	28	66.7	763	6	US-10-873-528-63	Sequence 63, Appl
253	28	66.7	116	7	US-11-050-857-397	Sequence 397, App	326	28	66.7	791	7	US-11-096-568A-19857	Sequence 19857, A
254	28	66.7	120	7	US-11-072-512-2575	Sequence 2575, Ap	327	28	66.7	830	6	US-10-995-561-899	Sequence 899, App
255	28	66.7	126	6	US-10-510-386-134	Sequence 134, App	328	28	66.7	830	6	US-11-096-568A-19856	Sequence 19856, A
256	28	66.7	140	7	US-11-075-046-44	Sequence 44, Appl	329	28	66.7	880	6	US-11-051-035-339	Sequence 339, App
257	28	66.7	151	7	US-11-072-512-3804	Sequence 3804, Ap	330	28	66.7	899	7	US-11-124-368A-290	Sequence 290, App
258	28	66.7	158	7	US-11-087-099-7941	Sequence 7941, Ap	331	28	66.7	912	6	US-10-493-909-75	Sequence 75, Appl
259	28	66.7	171	6	US-10-793-626-1656	Sequence 1656, Ap	332	28	66.7	917	6	US-10-493-909-87	Sequence 87, Appl
260	28	66.7	181	6	US-11-096-568A-33095	Sequence 33095, A	333	28	66.7	921	7	US-11-183-624-2	Sequence 2, Appli
261	28	66.7	209	7	US-11-096-568A-17701	Sequence 17701, A	334	28	66.7	923	7	US-11-052-554A-147	Sequence 147, App
262	28	66.7	210	7	US-11-052-554A-328	Sequence 328, App	335	28	66.7	924	6	US-10-857-780-20	Sequence 20, Appl
263	28	66.7	220	7	US-11-096-568A-17700	Sequence 17700, A	336	28	66.7	924	6	US-10-493-909-67	Sequence 67, Appl
264	28	66.7	222	7	US-11-069-642-119	Sequence 119, App	337	28	66.7	924	6	US-11-107-028-26	Sequence 26, Appl
265	28	66.7	224	7	US-11-096-568A-26103	Sequence 26103, A	338	28	66.7	949	7	US-11-052-554A-6	Sequence 6, Appli
266	28	66.7	246	7	US-11-087-099-9604	Sequence 9604, Ap	339	28	66.7	1011	7	US-11-069-642-111	Sequence 111, App
267	28	66.7	247	7	US-11-087-099-12017	Sequence 12017, A	340	28	66.7	1121	7	US-11-113-751-24	Sequence 24, Appl
268	28	66.7	248	7	US-11-150-883-21	Sequence 21, Appl	341	28	66.7	1121	7	US-11-113-751-24	Sequence 24, Appl
269	28	66.7	248	7	US-11-150-887-14	Sequence 14, Appl	342	28	66.7	1159	7	US-11-113-751-27	Sequence 27, Appl
270	28	66.7	248	7	US-11-241-035-28	Sequence 28, Appl	343	28	66.7	1242	6	US-10-330-773-412	Sequence 412, App
271	28	66.7	249	7	US-11-096-568A-17699	Sequence 17699, A	344	28	66.7	1242	6	US-10-770-726-46	Sequence 46, Appl
272	28	66.7	250	7	US-11-096-568A-693	Sequence 693, App	345	28	66.7	1271	6	US-10-330-773-415	Sequence 415, App
273	28	66.7	251	7	US-11-096-568A-18119	Sequence 18119, A	346	28	66.7	1296	6	US-11-124-368A-291	Sequence 291, App
274	28	66.7	257	7	US-11-096-568A-24246	Sequence 24246, A	347	28	66.7	1439	7	US-10-821-234-1102	Sequence 1102, Ap
275	28	66.7	262	7	US-11-087-099-9170	Sequence 9170, Ap	348	28	66.7	1452	6	US-11-052-554A-137	Sequence 137, App
276	28	66.7	278	7	US-11-096-568A-20244	Sequence 20244, A	349	28	66.7	1660	7	US-11-087-099-9594	Sequence 9594, Ap
277	28	66.7	293	6	US-10-878-586A-153	Sequence 153, App	350	28	66.7	2117	7	US-11-044-111-23	Sequence 23, Appl
278	28	66.7	293	7	US-11-096-568A-24843	Sequence 24843, A	351	28	66.7	3375	7	US-11-044-111-23	Sequence 23, Appl
279	28	66.7	297	7	US-11-052-554A-190	Sequence 190, App	352	28	66.7	3570	6	US-10-453-372-178	Sequence 178, App
280	28	66.7	301	7	US-11-096-568A-17636	Sequence 17636, A	353	28	66.7	3570	6	US-10-453-372-198	Sequence 198, App
281	28	66.7	318	7	US-11-096-568A-691	Sequence 691, App	354	28	66.7	3570	6	US-10-453-372-202	Sequence 202, App
282	28	66.7	320	7	US-11-096-568A-17635	Sequence 17635, A	355	28	66.7	3570	6	US-10-453-372-204	Sequence 204, App
283	28	66.7	323	7	US-11-096-568A-17634	Sequence 17634, A	356	28	66.7	3570	6	US-10-453-372-206	Sequence 206, App
284	28	66.7	329	7	US-11-123-402-28	Sequence 28, Appl	357	28	66.7	3570	6	US-10-503-575-130	Sequence 130, App
285	28	66.7	330	7	US-11-207-078-322	Sequence 322, App	358	27	64.3	19	6	US-11-154-227-69	Sequence 69, Appl
286	28	66.7	334	7	US-11-087-099-12010	Sequence 12010, A	359	27	64.3	43	7	US-11-043-788-504	Sequence 504, App
287	28	66.7	340	7	US-11-096-568A-20615	Sequence 20615, A	360	27	64.3	47	6	US-10-514-057-1	Sequence 1, Appli
288	28	66.7	344	7	US-11-087-099-4097	Sequence 4097, Ap	361	27	64.3	66	6	US-10-846-479-3	Sequence 11173, A
289	28	66.7	344	7	US-11-096-568A-18460	Sequence 18460, A	362	27	64.3	69	6	US-11-096-568A-11173	Sequence 11173, A
290	28	66.7	354	7	US-11-096-568A-18459	Sequence 18459, A	363	27	64.3	71	7	US-11-096-568A-927	Sequence 927, App
291	28	66.7	360	7	US-11-096-568A-20614	Sequence 20614, A	364	27	64.3	93	7	US-11-096-568A-275	Sequence 275, App
292	28	66.7	368	7	US-11-085-775-3	Sequence 3, Appli	365	27	64.3	96	6	US-10-993-543-275	Sequence 275, App
293	28	66.7	395	7	US-11-072-512-3413	Sequence 3413, Ap	366	27	64.3	99	7	US-11-084-554-201	Sequence 201, App
294	28	66.7	434	7	US-11-052-554A-167	Sequence 167, App	367	27	64.3	99	7	US-11-136-250-201	Sequence 201, App
295	28	66.7	437	7	US-11-087-099-11532	Sequence 11532, A	368	27	64.3	101	7	US-11-055-822-272	Sequence 272, App
296	28	66.7	444	7	US-11-087-099-9103	Sequence 9103, Ap	369	27	64.3	102	7	US-11-096-568A-24201	Sequence 24201, A
297	28	66.7	457	7	US-11-055-822-80	Sequence 80, Appl	370	27	64.3	108	7	US-11-064-174-68	Sequence 926, App
298	28	66.7	458	6	US-10-618-320A-1	Sequence 1, Appli	371	27	64.3	108	7	US-11-096-568A-926	Sequence 926, App
299	28	66.7	459	7	US-11-200-486-2	Sequence 2, Appli	372	27	64.3	110	6	US-10-982-440-34	Sequence 34, Appl
300	28	66.7	463	7	US-11-087-099-7935	Sequence 7935, Ap	373	27	64.3	110	7	US-11-049-536-256	Sequence 256, App
301	28	66.7	493	7	US-11-096-568A-25638	Sequence 25638, A	374	27	64.3	114	7	US-11-049-536-548	Sequence 548, App
302	28	66.7	503	7	US-11-078-991-1	Sequence 1, Appli	375	27	64.3	119	5	US-11-096-568A-925	Sequence 925, App
303	28	66.7	503	7	US-11-072-512-3266	Sequence 3266, Ap	376	27	64.3	121	6	US-09-978-360A-344	Sequence 344, Ap
304	28	66.7	513	7	US-11-096-568A-25637	Sequence 25637, A	377	27	64.3	121	6	US-10-821-234-1343	Sequence 1343, Ap
305	28	66.7	522	7	US-11-010-239-52	Sequence 52, Appl	378	27	64.3	123	6	US-10-993-543-10	Sequence 10, Appl
306	28	66.7	523	6	US-10-880-881-18	Sequence 18, Appl	379	27	64.3	123	6	US-10-993-543-34	Sequence 34, Appl
307	28	66.7	546	7	US-11-143-980-38	Sequence 38, Appl	380	27	64.3	123	6	US-10-993-543-142	Sequence 142, App
308	28	66.7	556	7	US-11-087-099-3621	Sequence 3621, Ap	381	27	64.3	123	6	US-11-072-512-2295	Sequence 2295, Ap
309	28	66.7	574	6	US-10-507-275-7	Sequence 7, Appli	382	27	64.3	126	7	US-11-245-689-6	Sequence 6, Appli
310	28	66.7	574	6	US-10-770-726-50	Sequence 50, Appl	383	27	64.3	126	7	US-11-245-689-4	Sequence 4, Appli
311	28	66.7	577	7	US-11-200-486-6	Sequence 6, Appli	384	27	64.3	126	7	US-11-245-689-27	Sequence 27, Appl
312	28	66.7	577	7	US-11-096-568A-25636	Sequence 25636, A	385	27	64.3	129	7	US-11-245-689-23	Sequence 23, Appl
313	28	66.7	593	7	US-11-040-488-2	Sequence 2, Appli	386	27	64.3	132	7	US-11-245-689-28	Sequence 28, Appl
314	28	66.7	618	6	US-10-501-035-368	Sequence 368, App	387	27	64.3	135	7	US-11-245-689-2	Sequence 2, Appli
315	28	66.7	618	7	US-11-052-554A-150	Sequence 150, App	388	27	64.3	138	7	US-11-072-512-3246	Sequence 3246, Ap
316	28	66.7	635	6	US-10-523-503-38	Sequence 38, Appl	389	27	64.3	139	7		
317	28	66.7	639	7	US-11-052-554A-165	Sequence 165, App	390	27	64.3				

391	27	64.3	141	7	US-11-245-689-25	Sequence 25, Appl	464	27	64.3	285	7	US-11-096-568A-7886	Sequence 7886, Ap
392	27	64.3	144	7	US-11-245-689-13	Sequence 13, Appl	465	27	64.3	285	7	US-11-096-568A-34122	Sequence 34122, A
393	27	64.3	144	7	US-11-096-568A-18743	Sequence 18743, A	466	27	64.3	292	6	US-10-967-527A-19	Sequence 19, Appl
394	27	64.3	150	7	US-11-245-689-1	Sequence 1, Appl	467	27	64.3	292	7	US-11-096-568A-7885	Sequence 7885, Ap
395	27	64.3	150	7	US-11-245-689-10	Sequence 10, Appl	468	27	64.3	293	6	US-10-742-634-7	Sequence 7, Appl
396	27	64.3	156	7	US-11-245-689-24	Sequence 24, Appl	469	27	64.3	293	7	US-11-221-849-2	Sequence 2, Appl
397	27	64.3	159	7	US-11-096-568A-8681	Sequence 8681, Ap	470	27	64.3	293	7	US-11-242-294-2	Sequence 2, Appl
398	27	64.3	163	7	US-11-096-568A-8680	Sequence 8680, Ap	471	27	64.3	296	7	US-11-096-051-18	Sequence 18, Appl
399	27	64.3	163	7	US-11-096-568A-23491	Sequence 23491, A	472	27	64.3	301	7	US-11-096-568A-7884	Sequence 7884, Ap
400	27	64.3	166	7	US-11-245-689-26	Sequence 26, Appl	473	27	64.3	302	6	US-10-453-372-780	Sequence 780, App
401	27	64.3	166	7	US-11-096-568A-24434	Sequence 24434, A	474	27	64.3	302	6	US-10-453-372-782	Sequence 782, App
402	27	64.3	176	7	US-11-096-568A-8679	Sequence 8679, Ap	475	27	64.3	302	6	US-10-453-372-788	Sequence 788, App
403	27	64.3	177	7	US-11-245-689-17	Sequence 17, Appl	476	27	64.3	302	6	US-10-453-372-792	Sequence 792, App
404	27	64.3	180	7	US-11-245-689-18	Sequence 18, Appl	477	27	64.3	305	6	US-10-714-887-354	Sequence 354, App
405	27	64.3	183	7	US-11-096-568A-24433	Sequence 24433, A	478	27	64.3	305	6	US-10-712-512-2670	Sequence 2670, Ap
406	27	64.3	185	7	US-11-072-512-1972	Sequence 1972, Ap	479	27	64.3	307	7	US-11-096-568A-31301	Sequence 31301, A
407	27	64.3	186	7	US-11-245-689-15	Sequence 15, Appl	480	27	64.3	309	6	US-10-453-372-778	Sequence 778, App
408	27	64.3	188	7	US-11-054-515-3235	Sequence 3235, Ap	481	27	64.3	311	7	US-11-096-568A-11284	Sequence 11284, A
409	27	64.3	189	7	US-11-087-099-8115	Sequence 8115, Ap	482	27	64.3	313	7	US-11-245-689-44	Sequence 44, Appl
410	27	64.3	190	6	US-10-131-826A-158	Sequence 158, App	483	27	64.3	314	7	US-11-096-568A-21558	Sequence 21558, A
411	27	64.3	190	6	US-10-973-115B-158	Sequence 158, App	484	27	64.3	318	7	US-11-087-099-12185	Sequence 12185, A
412	27	64.3	196	7	US-11-096-568A-10176	Sequence 10176, A	485	27	64.3	321	7	US-11-096-568A-1203	Sequence 1203, Ap
413	27	64.3	196	7	US-11-096-568A-23804	Sequence 23804, A	486	27	64.3	324	6	US-10-698-618-2	Sequence 2, Appl
414	27	64.3	207	7	US-11-054-515-3234	Sequence 3234, Ap	487	27	64.3	326	7	US-11-096-568A-12080	Sequence 12080, A
415	27	64.3	217	7	US-11-087-099-4366	Sequence 4366, Ap	488	27	64.3	327	7	US-11-096-568A-16583	Sequence 16583, A
416	27	64.3	217	7	US-11-054-515-3233	Sequence 3233, Ap	489	27	64.3	328	7	US-11-096-568A-21557	Sequence 21557, A
417	27	64.3	224	7	US-11-096-568A-10783	Sequence 10783, A	490	27	64.3	331	7	US-11-096-568A-16148	Sequence 16148, A
418	27	64.3	230	7	US-11-096-568A-8424	Sequence 8424, Ap	491	27	64.3	333	7	US-10-514-057-6	Sequence 6, Appl
419	27	64.3	231	7	US-11-096-568A-25374	Sequence 25374, A	492	27	64.3	336	6	US-11-096-568A-14878	Sequence 14878, A
420	27	64.3	232	7	US-11-096-568A-14658	Sequence 14658, A	493	27	64.3	336	7	US-11-096-568A-21079	Sequence 21079, A
421	27	64.3	234	7	US-11-096-568A-10782	Sequence 10782, A	494	27	64.3	336	7	US-11-055-822-270	Sequence 270, App
422	27	64.3	239	7	US-11-054-515-3232	Sequence 3232, Ap	495	27	64.3	340	7	US-11-072-512-3319	Sequence 3319, Ap
423	27	64.3	240	7	US-11-087-099-8001	Sequence 8001, Ap	496	27	64.3	340	7	US-11-087-099-8225	Sequence 8225, Ap
424	27	64.3	241	7	US-11-054-515-2055	Sequence 2055, Ap	497	27	64.3	342	7	US-11-096-568A-31300	Sequence 31300, A
425	27	64.3	243	7	US-11-096-568A-34123	Sequence 34123, A	498	27	64.3	344	6	US-10-967-527A-24	Sequence 24, Appl
426	27	64.3	245	7	US-11-096-568A-32632	Sequence 32632, A	499	27	64.3	346	7	US-11-096-568A-1201	Sequence 1201, Ap
427	27	64.3	247	6	US-10-514-040-2	Sequence 2, Appl	500	27	64.3				
428	27	64.3	247	6	US-10-296-865-2	Sequence 4, Appl							
429	27	64.3	247	6	US-10-286-865-4	Sequence 4, Appl							
430	27	64.3	247	6	US-11-054-515-1685	Sequence 1685, Ap							
431	27	64.3	249	6	US-10-967-527A-21	Sequence 21, Appl							
432	27	64.3	249	6	US-11-054-515-2065	Sequence 2065, Ap							
433	27	64.3	251	7	US-11-054-515-917	Sequence 917, App							
434	27	64.3	251	7	US-11-054-515-925	Sequence 925, App							
435	27	64.3	251	7	US-11-054-515-951	Sequence 951, App							
436	27	64.3	251	7	US-11-054-515-1496	Sequence 1496, Ap							
437	27	64.3	252	7	US-11-054-515-1583	Sequence 1583, Ap							
438	27	64.3	253	7	US-11-054-515-1249	Sequence 1249, Ap							
439	27	64.3	253	7	US-11-054-515-1364	Sequence 1364, Ap							
440	27	64.3	254	7	US-11-054-515-1544	Sequence 1544, Ap							
441	27	64.3	255	7	US-11-054-515-1233	Sequence 1233, Ap							
442	27	64.3	255	7	US-11-054-515-1626	Sequence 1626, Ap							
443	27	64.3	256	7	US-11-054-515-1558	Sequence 1558, Ap							
444	27	64.3	256	7	US-11-054-515-1559	Sequence 1559, Ap							
445	27	64.3	256	7	US-11-054-515-1791	Sequence 1791, Ap							
446	27	64.3	257	7	US-11-096-051-12	Sequence 12, Appl							
447	27	64.3	258	7	US-11-054-515-862	Sequence 862, App							
448	27	64.3	258	7	US-11-054-515-1173	Sequence 1173, Ap							
449	27	64.3	258	7	US-11-054-515-3246	Sequence 3246, Ap							
450	27	64.3	259	7	US-11-054-515-1247	Sequence 1247, Ap							
451	27	64.3	259	7	US-11-140-416-33	Sequence 33, Appl							
452	27	64.3	261	7	US-11-207-078-293	Sequence 293, App							
453	27	64.3	262	6	US-10-793-826-3158	Sequence 3158, Ap							
454	27	64.3	263	7	US-11-072-512-3517	Sequence 3517, Ap							
455	27	64.3	264	7	US-11-096-568A-16150	Sequence 16150, A							
456	27	64.3	266	6	US-10-131-826A-428	Sequence 428, App							
457	27	64.3	266	6	US-10-973-115B-428	Sequence 428, App							
458	27	64.3	266	7	US-11-096-568A-23193	Sequence 23193, A							
459	27	64.3	270	7	US-11-087-099-9342	Sequence 9342, Ap							
460	27	64.3	271	7	US-11-096-568A-8423	Sequence 8423, Ap							
461	27	64.3	276	7	US-11-087-099-4138	Sequence 4138, Ap							
462	27	64.3	280	6	US-10-821-234-1307	Sequence 1307, Ap							
463	27	64.3	282	7	US-11-096-568A-14879	Sequence 14879, A							

ALIGNMENTS

RESULT 1

US-11-024-959-335
; Sequence 335, Application US/11024959
; Publication No. US20060010516A1
; GENERAL INFORMATION:
; APPLICANT: FORSTER, RICHARD L.
; APPLICANT: CONNETT, MARIE B.
; APPLICANT: EMERSON, SARAH JANE
; APPLICANT: GRIGOR, MURRAY ROBERT
; APPLICANT: HIGGINS, COLLEEN M.
; APPLICANT: LUND, STEVEN TROY
; APPLICANT: MAGUSIN, ANDREAS
; APPLICANT: KODRZYCKI, BOB
; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS
; FILE REFERENCE: 044463-0360
; CURRENT APPLICATION NUMBER: US/11/024,959
; CURRENT FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: 60/533,036
; PRIOR FILING DATE: 2003-12-30
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 335
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Eucalyptus sp.
US-11-024-959-335

Query Match 90.5%; Score 38; DB 7; Length 317;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 GAEGSPG 7
Db      311 GAEGSPG 317

RESULT 2
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radote T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Query Match      90.5%; Score 38; DB 7; Length 1464;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      1019 GAEGSPG 1025

RESULT 3
US-11-052-554A-148
; Sequence 148, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3

Qy      1 GAEGSPG 7
Db      1019 GAEGSPG 1025

Query Match      90.5%; Score 38; DB 7; Length 1464;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      1019 GAEGSPG 1025

Query Match      90.5%; Score 38; DB 7; Length 1464;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; SEQ ID NO 148
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-148

Query Match      85.7%; Score 36; DB 7; Length 749;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      740 GAEGSPGL 747

RESULT 4
US-11-051-720-1372
; Sequence 1372, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051,720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1372
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1372

Query Match      85.7%; Score 36; DB 7; Length 1081;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 GAEGSPGL 8
Db      1036 GAQGAFL 1043

RESULT 5
US-11-182-016-5
; Sequence 5, Application US/11182016
; Publication No. US20060019294A1
; GENERAL INFORMATION:
; APPLICANT: SUGEN, INC.
; TITLE OF INVENTION: TYROSINE KINASE SUBSTRATE (TKS) PROTEINS
; FILE REFERENCE: 038602/0102
; CURRENT APPLICATION NUMBER: US/11/182,016
; CURRENT FILING DATE: 2005-07-15
; PRIOR APPLICATION NUMBER: US/09/958,359
; PRIOR FILING DATE: 2002-02-05
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1519
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Tks 107
US-11-182-016-5

Query Match      85.7%; Score 36; DB 7; Length 1519;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 AEGSPGL 8
Db      259 AEGSPGL 265
```

```
RESULT 6
US-10-995-561-911
; Sequence 911, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 911
; LENGTH: 1767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-911
Query Match      85.7%; Score 36; DB 6; Length 1767;
Best Local Similarity 75.0%; Pred. No. 97;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      ||:||:|
Db      997 GAQGAPGL 1004

RESULT 7
US-10-995-561-914
; Sequence 914, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 914
; LENGTH: 1767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-914
Query Match      85.7%; Score 36; DB 6; Length 1767;
Best Local Similarity 75.0%; Pred. No. 97;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      ||:||:|
Db      997 GAQGAPGL 1004

RESULT 8
US-10-995-561-912
; Sequence 912, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 912
```

```
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-912
Query Match      85.7%; Score 36; DB 6; Length 1806;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      ||:||:|
Db      1036 GAQGAPGL 1043

RESULT 9
US-10-995-561-915
; Sequence 915, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-915
Query Match      85.7%; Score 36; DB 6; Length 1806;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      ||:||:|
Db      1036 GAQGAPGL 1043

RESULT 10
US-11-051-720-1446
; Sequence 1446, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051,720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1446
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1446
Query Match      85.7%; Score 36; DB 7; Length 1806;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      ||:||:|
Db      1036 GAQGAPGL 1043

RESULT 11
US-11-051-720-1447
; Sequence 1447, Application US/11051720
```

```
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051.720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1447
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1447

Query Match      85.7%; Score 36; DB 7; Length 1806;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      ||:|:|
Db      1036 GAQGAPGL 1043

RESULT 12
US-10-995-561-910
; Sequence 910, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 910
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-910

Query Match      85.7%; Score 36; DB 6; Length 1818;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      ||:|:|
Db      1048 GAQGAPGL 1055

RESULT 13
US-10-995-561-913
; Sequence 913, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 913
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-913
```

```
Query Match      85.7%; Score 36; DB 6; Length 1818;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      ||:|:|
Db      1048 GAQGAPGL 1055

RESULT 14
US-11-096-568A-11972
; Sequence 11972, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 11972
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(150)
; OTHER INFORMATION: Ceres Seq. ID no. 13659646
US-11-096-568A-11972

Query Match      83.3%; Score 35; DB 7; Length 150;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      ||:|:|
Db      84 GAQGDPGL 91

RESULT 15
US-10-485-517-354
; Sequence 354, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 354
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-354

Query Match      83.3%; Score 35; DB 6; Length 260;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPGL 8
      ||:|:|
Db      146 GVEGNPGL 153
```

RESULT 16

US-10-485-517-150
; Sequence 150, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 150
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-150

Query Match 83.3%; Score 35; DB 6; Length 261;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 147 GVEGNPGL 154

RESULT 17

US-10-821-234-964
; Sequence 964, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: Pt_SEQ_genes Version 1.0
; SEQ ID NO 964
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-964

Query Match 83.3%; Score 35; DB 6; Length 1166;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 873 GSEGSFG 879

RESULT 18

US-11-052-554A-139
; Sequence 139, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 139
; LENGTH: 1306
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37RV
US-11-052-554A-139

Query Match 83.3%; Score 35; DB 7; Length 1306;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 118 GADGSPG 124

RESULT 19

US-10-501-035-331
; Sequence 331, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 331
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-331

Query Match 83.3%; Score 35; DB 6; Length 1464;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7

Db 752 GADGSPG 758

RESULT 20

US-11-186-284-28
; Sequence 28, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER


```
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-28

Query Match      83.3%; Score 35; DB 7; Length 1464;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      752 GADGSPG 758

RESULT 21
US-11-021-603-2
; Sequence 2, Application US/11021603
; Publication No. US2006003954A1
; GENERAL INFORMATION:
; APPLICANT: Berli, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254U1
; CURRENT APPLICATION NUMBER: US/11/021,603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149,352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-021-603-2

Query Match      83.3%; Score 35; DB 7; Length 1464;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      752 GADGSPG 758

RESULT 22
US-11-186-284-33
; Sequence 33, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.

; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-33

Query Match      83.3%; Score 35; DB 7; Length 1466;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      1173 GSEGSPG 1179

RESULT 23
US-10-821-234-1096
; Sequence 1096, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1096
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1096

Query Match      83.3%; Score 35; DB 6; Length 1467;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      755 GADGSPG 761

RESULT 24
US-11-193-205-1
; Sequence 1, Application US/11193205
; Publication No. US20060024285A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Y-Ming A.
; TITLE OF INVENTION: Carcinogen Detoxification Composition
; TITLE OF INVENTION: Method
; FILE REFERENCE: 049202/295647
```

```
; CURRENT APPLICATION NUMBER: US/11/193,205
; CURRENT FILING DATE: 2005-07-29
; PRIOR APPLICATION NUMBER: EP 04 018113.3
; PRIOR FILING DATE: 2004-07-30
; PRIOR APPLICATION NUMBER: 60/600,367
; PRIOR FILING DATE: 2004-08-11
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-193-205-1

Query Match      81.0%; Score 34; DB 7; Length 295;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      231 GQDQSPGL 238

RESULT 25
US-10-995-561-532
; Sequence 532, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 532
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-532

Query Match      81.0%; Score 34; DB 6; Length 520;
Best Local Similarity 75.0%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      376 GEQSPGL 383

RESULT 26
US-11-072-512-3814
; Sequence 3814, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI

; CURRENT APPLICATION NUMBER: US/11/193,205
; CURRENT FILING DATE: 2005-07-29
; PRIOR APPLICATION NUMBER: EP 04 018113.3
; PRIOR FILING DATE: 2004-07-30
; PRIOR APPLICATION NUMBER: 60/600,367
; PRIOR FILING DATE: 2004-08-11
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-193-205-1

; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cdna
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3814
; LENGTH: 571
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3814

Query Match      81.0%; Score 34; DB 7; Length 571;
Best Local Similarity 87.5%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      11 GAEGSPKL 18

RESULT 27
US-10-821-234-1481
; Sequence 1481, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1481
; LENGTH: 618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1481

Query Match      81.0%; Score 34; DB 6; Length 618;
Best Local Similarity 85.7%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      106 GAEGKPG 112

RESULT 28
US-10-330-773-700
; Sequence 700, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
```

```
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-330-773-700

Query Match      81.0%; Score 34; DB 6; Length 822;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGSPG 7
   ||:||||
Db 63 GAKGSPG 69

RESULT 29
US-11-135-855-30
; Sequence 30, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-30

Query Match      78.6%; Score 33; DB 7; Length 288;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 GAGSPG 7
   ||:||||
Db 47 GAKGSPG 53

RESULT 30
US-11-258-647-15
; Sequence 15, Application US/11258647
; Publication No. US20060040360A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; TITLE OF INVENTION: HOMOLOG ZACRP7
; FILE REFERENCE: 99-31C2
; CURRENT APPLICATION NUMBER: US/11/258,647
; CURRENT FILING DATE: 2005-10-25
; PRIOR APPLICATION NUMBER: US 10/234,000
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US 09/577,298
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: US 60/158,448
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: US 60/145,589
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/136,289
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Mus musculus
US-11-258-647-15

Query Match      78.6%; Score 33; DB 7; Length 289;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 GAGSPG 7
   ||:||||
Db 47 GAKGSPG 53

RESULT 31
US-11-135-855-31
; Sequence 31, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-31

Query Match      78.6%; Score 33; DB 7; Length 303;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 GAGSPG 7
   ||:||||
Db 61 GAKGSPG 67

RESULT 32
US-11-258-647-2
; Sequence 2, Application US/11258647
; Publication No. US20060040360A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; TITLE OF INVENTION: HOMOLOG ZACRP7
; FILE REFERENCE: 99-31C2
; CURRENT APPLICATION NUMBER: US/11/258,647
; CURRENT FILING DATE: 2005-10-25
; PRIOR APPLICATION NUMBER: US 10/234,000
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US 09/577,298
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: US 60/158,448
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: US 60/145,589
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/136,289
```

```
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-258-647-2

Query Match      78.6%; Score 33; DB 7; Length 303;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
      |||||
Db      61 GAGSPG 67

RESULT 33
US-11-072-512-2505
; Sequence 2505, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2505
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-2505

Query Match      78.6%; Score 33; DB 7; Length 756;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 AEGSPGL 8
      |:|||||
Db      475 ADGSPGL 481

RESULT 34
US-10-995-561-983
; Sequence 983, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 983
; LENGTH: 828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-983

Query Match      78.6%; Score 33; DB 6; Length 828;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      400 GAGSPGV 407

RESULT 35
US-10-995-561-981
; Sequence 981, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 981
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-981

Query Match      78.6%; Score 33; DB 6; Length 918;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
      |||||
Db      400 GAGSPGV 407

RESULT 36
US-10-995-561-982
; Sequence 982, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 982
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-982

Query Match      78.6%; Score 33; DB 6; Length 1019;
Best Local Similarity 75.0%; Pred. No. 2e+02;
```

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 400 GAGSPGV 407

RESULT 37

US-11-096-070-34
; Sequence 34, Application US/11096070
; Publication No. US20050287098A1
; GENERAL INFORMATION:
; APPLICANT: SUN, TUNG-TIEN
; APPLICANT: CAO, QIONG
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH
; FILE REFERENCE: 71369.274US2
; CURRENT APPLICATION NUMBER: US/11/096,070
; CURRENT FILING DATE: 2005-03-31
; PRIOR APPLICATION NUMBER: 60/558,341
; PRIOR FILING DATE: 2004-03-31
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 3.3
; SEQ ID NO 34
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Rattus sp.
US-11-096-070-34

Query Match 76.2%; Score 32; DB 7; Length 117;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8

Db 22 GSDGQFGL 29

RESULT 38

US-11-087-227-18
; Sequence 18, Application US/11087227
; Publication No. US20050260566A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-18

Query Match 76.2%; Score 32; DB 7; Length 164;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AEGSPG 7

Db 127 AEGSPG 132

RESULT 39

US-11-087-227-20
; Sequence 20, Application US/11087227

; Publication No. US20050260566A1
; GENERAL INFORMATION:

; APPLICANT: Fischer, Timothy J.
; APPLICANT: Malinowski, Douglas P.
; APPLICANT: Taylor, Adriann J.
; APPLICANT: Parker, Margaret R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF CERVICAL DISEASE
; FILE REFERENCE: 046143/287139
; CURRENT APPLICATION NUMBER: US/11/087,227
; CURRENT FILING DATE: 2005-03-23
; PRIOR APPLICATION NUMBER: 60/556,495
; PRIOR FILING DATE: 2004-03-24
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 164
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-087-227-20

Query Match 76.2%; Score 32; DB 7; Length 164;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 AEGSPG 7

Db 127 AEGSPG 132

RESULT 40

US-11-197-133A-62
; Sequence 62, Application US/11197133A
; Publication No. US20060040361A1
; GENERAL INFORMATION:
; APPLICANT: De Le Fuente Jose de Jesus
; APPLICANT: Kocan Katherine M.
; APPLICANT: Garcia-Almazan Consuelo
; APPLICANT: Blouin Edwin F.
; TITLE OF INVENTION: Protective antigens and vaccines for the control of multi species-
; TITLE OF INVENTION: Infestations
; FILE REFERENCE: 57338/05-261
; CURRENT APPLICATION NUMBER: US/11/197,133A
; CURRENT FILING DATE: 2005-08-04
; PRIOR APPLICATION NUMBER: US 10/972789
; PRIOR FILING DATE: 2004-10-25
; PRIOR APPLICATION NUMBER: US 10/425563
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/376251
; PRIOR FILING DATE: 2002-04-29
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 62
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Haemaphysalis punctata
US-11-197-133A-62

Query Match 76.2%; Score 32; DB 7; Length 179;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSP 6

Db 83 GAEGSP 88

RESULT 41

US-10-131-826A-316
; Sequence 316, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

```
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 316
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-316

Query Match 76.2%; Score 32; DB 6; Length 288;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 48 GAGSPG 54

RESULT 42
US-10-973-115B-316
; Sequence 316, Application US/10973115B
; Publication No. US20060040351A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Deanoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Gurney, Austin L.
```

```
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 39870-3330R1C300C1
; CURRENT APPLICATION NUMBER: US/10/973,115B
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 10/145,747
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 10/028,072
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/581,742
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: PCT/US00/05746
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/135,736
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 60/123,090
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 316
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-973-115B-316

Query Match 76.2%; Score 32; DB 6; Length 288;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 48 GAGSPG 54

RESULT 43
US-10-995-561-999
; Sequence 999, Application US/10995561
; Publication No. US2005027054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CU001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 999
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-999

Query Match 76.2%; Score 32; DB 6; Length 295;
Best Local Similarity 85.7%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
Db 155 GAGSPG 161

RESULT 44
US-11-177-506-37
```

Mon Mar 13 10:53:07 2006

; Sequence 37, Application US/11177506
; Publication No. US20060029956A1
; GENERAL INFORMATION:
; APPLICANT: Beyer, Wayne F.
; APPLICANT: Venetta, Thomas M.
; APPLICANT: Groelke, John W.
; APPLICANT: Blaesus, Rainer H.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
; TITLE OF INVENTION: DETECTION OF OVARIAN DISEASE
; FILE REFERENCE: 46143/294851
; CURRENT APPLICATION NUMBER: US/11/177,506
; CURRENT FILING DATE: 2005-07-08
; PRIOR APPLICATION NUMBER: 60/586,856
; PRIOR FILING DATE: 2004-07-09
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-177-506-37

Query Match 76.2%; Score 32; DB 7; Length 299;
Best Local Similarity 85.7%; Pred. NO. 87;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
|||
Db 48 GAGGSPG 54

RESULT 45
US-11-087-099-12163
; Sequence 12163, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 12163
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Caulobacter crescentus CB15
US-11-087-099-12163

Query Match 76.2%; Score 32; DB 7; Length 350;
Best Local Similarity 75.0%; Pred. NO. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|||
Db 266 GAGGKPL 273

RESULT 46
US-11-072-512-3291
; Sequence 3291, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI

; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3291
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3291

Query Match 76.2%; Score 32; DB 7; Length 406;
Best Local Similarity 85.7%; Pred. NO. 1.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
|||
Db 339 GAGGSPG 345

RESULT 47
US-11-072-512-3166
; Sequence 3166, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: 084335-0191
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3166
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3166

Query Match 76.2%; Score 32; DB 7; Length 438;
Best Local Similarity 85.7%; Pred. NO. 1.3e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
||| |||

Db 298 GAAGSPG 304

RESULT 48

US-11-096-070-4

; Sequence 4, Application US/11096070

; Publication No. US20050287098A1

; GENERAL INFORMATION:

; APPLICANT: CAO, QIONG

; APPLICANT: SUN, TUNG-TIEN

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH

; FILE REFERENCE: 71369.274US2

; CURRENT APPLICATION NUMBER: US/11/096,070

; CURRENT FILING DATE: 2005-03-31

; PRIOR APPLICATION NUMBER: 60/558,341

; PRIOR FILING DATE: 2004-03-31

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 4

; LENGTH: 531

; TYPE: PRT

; ORGANISM: Rattus sp.

US-11-096-070-4

Query Match

Best Local Similarity 76.2%; Score 32; DB 7; Length 531;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|::|||

Db 139 GSDGQPGL 146

RESULT 49

US-11-096-070-2

; Sequence 2, Application US/11096070

; Publication No. US20050287098A1

; GENERAL INFORMATION:

; APPLICANT: SUN, TUNG-TIEN

; APPLICANT: CAO, QIONG

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH

; FILE REFERENCE: 71369.274US2

; CURRENT APPLICATION NUMBER: US/11/096,070

; CURRENT FILING DATE: 2005-03-31

; PRIOR APPLICATION NUMBER: 60/558,341

; PRIOR FILING DATE: 2004-03-31

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 2

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Rattus sp.

US-11-096-070-2

Query Match

Best Local Similarity 76.2%; Score 32; DB 7; Length 549;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|::|||

Db 157 GSDGQPGL 164

RESULT 50

US-11-096-070-6

; Sequence 6, Application US/11096070

; Publication No. US20050287098A1

; GENERAL INFORMATION:

; APPLICANT: SUN, TUNG-TIEN

; APPLICANT: CAO, QIONG

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH

; FILE REFERENCE: 71369.274US2

; CURRENT APPLICATION NUMBER: US/11/096,070

; CURRENT FILING DATE: 2005-03-31

; PRIOR APPLICATION NUMBER: 60/558,341

; PRIOR FILING DATE: 2004-03-31

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 6

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Rattus sp.

US-11-096-070-6

Query Match

Best Local Similarity 76.2%; Score 32; DB 7; Length 549;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|::|||

Db 157 GSDGQPGL 164

RESULT 51

US-11-096-070-8

; Sequence 8, Application US/11096070

; Publication No. US20050287098A1

; GENERAL INFORMATION:

; APPLICANT: SUN, TUNG-TIEN

; APPLICANT: CAO, QIONG

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH

; FILE REFERENCE: 71369.274US2

; CURRENT APPLICATION NUMBER: US/11/096,070

; CURRENT FILING DATE: 2005-03-31

; PRIOR APPLICATION NUMBER: 60/558,341

; PRIOR FILING DATE: 2004-03-31

; NUMBER OF SEQ ID NOS: 38

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 8

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Mus sp.

US-11-096-070-8

Query Match

Best Local Similarity 76.2%; Score 32; DB 7; Length 549;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
|::|||

Db 157 GSDGQPGL 164

RESULT 52

US-11-052-554A-164

; Sequence 164, Application US/11052554A

; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 164

; LENGTH: 576


```
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-164

Query Match      76.2%; Score 32; DB 7; Length 576;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      119 GAGGAPG 125

RESULT 53
US-11-052-554A-152
; Sequence 152, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 152
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-152

Query Match      76.2%; Score 32; DB 7; Length 615;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      541 GAGGAPGL 548

RESULT 54
US-10-501-035-254
; Sequence 254, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 254
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-254

Query Match      76.2%; Score 32; DB 6; Length 737;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GAEGSPG 7
Db      119 GAGGAPG 125

RESULT 55
US-11-052-554A-166
; Sequence 166, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166
; LENGTH: 801
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-166

Query Match      76.2%; Score 32; DB 7; Length 801;
Best Local Similarity 75.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      304 GAGGAPGL 311

RESULT 56
US-11-052-554A-159
; Sequence 159, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 159
; LENGTH: 837
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-159

Query Match      76.2%; Score 32; DB 7; Length 837;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GAEGSPGL 8
Db      393 GAGGAPGL 400

RESULT 57
US-11-186-641A-2
; Sequence 2, Application US/11186641A
; Publication No. US20060019301A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
```

```
; APPLICANT: Bjoernvad, Mads
; APPLICANT: Cherry, Joel
; APPLICANT: Jones, Aubrey
; APPLICANT: Fischer, Amanda
; TITLE OF INVENTION: Methods of producing mutant polynucleotides
; FILE REFERENCE: 10655-500-US
; CURRENT APPLICATION NUMBER: US/11/186.641A
; CURRENT FILING DATE: 2005-07-20
; PRIOR APPLICATION NUMBER: 60/589,502
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: 60/633,756
; PRIOR FILING DATE: 2004-12-06
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-11-186-641A-2

Query Match          76.2%; Score 32; DB 7; Length 861;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
   |||:|||
Db 752 GAGGNPGL 759

RESULT 58
US-11-052-554A-160
; Sequence 160, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 160
; LENGTH: 914
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-160

Query Match          76.2%; Score 32; DB 7; Length 914;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
   |||:|||
Db 433 GAGGAPGL 440

RESULT 59
US-10-821-234-1514
; Sequence 1514, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
```

```
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: Pct_SEQ_genes Version 1.0
; SEQ ID NO 1514
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1514

Query Match          76.2%; Score 32; DB 6; Length 915;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||:|||
Db 775 GAAGSPG 781

RESULT 60
US-10-995-561-1003
; Sequence 1003, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1003
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-1003

Query Match          76.2%; Score 32; DB 6; Length 915;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPG 7
   |||:|||
Db 775 GAAGSPG 781

RESULT 61
US-10-995-561-1000
; Sequence 1000, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1000
; LENGTH: 917
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-1000

Query Match          76.2%; Score 32; DB 6; Length 917;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 GAEGSPG 7
      |||||
Db      775 GAAGSPG 781

RESULT 62
US-10-995-561-1004
; Sequence 1004, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1004
; LENGTH: 940
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-1004

Query Match      76.2%; Score 32; DB 6; Length 940;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
      |||||
Db      775 GAAGSPG 781

RESULT 63
US-10-995-561-1001
; Sequence 1001, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1001
; LENGTH: 969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-1001

Query Match      76.2%; Score 32; DB 6; Length 969;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
      |||||
Db      829 GAAGSPG 835

RESULT 64
US-10-995-561-998
; Sequence 998, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559

QY      1 GAEGSPG 7
      |||||
Db      829 GAAGSPG 835

Query Match      76.2%; Score 32; DB 6; Length 971;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
      |||||
Db      829 GAAGSPG 835

RESULT 65
US-10-995-561-997
; Sequence 997, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 997
; LENGTH: 994
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-997

Query Match      76.2%; Score 32; DB 6; Length 994;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
      |||||
Db      829 GAAGSPG 835

RESULT 66
US-10-821-234-1431
; Sequence 1431, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_SEQ_genes Version 1.0
; SEQ ID NO 1431
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1431

Query Match      76.2%; Score 32; DB 6; Length 1366;
Best Local Similarity 75.0%; Pred. No. 4e+02;
```

```
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 910 GAVGSPGV 917

RESULT 67
US-11-186-284-31
; Sequence 31, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 1366
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-31

Query Match 76.2%; Score 32; DB 7; Length 1366;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 910 GAVGSPGV 917

RESULT 68
US-11-124-368A-329
; Sequence 329, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 2112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 329
; LENGTH: 1736
; TYPE: PRT

Query Match 76.2%; Score 32; DB 7; Length 1366;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAEGSPGL 8
Db 910 GAVGSPGV 917

RESULT 69
US-11-126-313-38
; Sequence 38, Application US/11126313
; Publication No. US20050288489A1
; GENERAL INFORMATION:
; APPLICANT: Hirsch, Joel
; TITLE OF INVENTION: VOLTAGE-DEPENDENT CALCIUM CHANNEL BETA SUBUNIT FUNCTIONAL CORE
; FILE REFERENCE: P-6758-US
; CURRENT APPLICATION NUMBER: US/11/126,313
; CURRENT FILING DATE: 2005-05-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 38
; LENGTH: 1783
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-11-126-313-38

Query Match 76.2%; Score 32; DB 7; Length 1783;
Best Local Similarity 100.0%; Pred. No. 5.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAEGSP 6
Db 1628 GAEGSP 1633

RESULT 70
US-11-096-568A-12579
; Sequence 12579, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 12579
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Triticum aestivum
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(76)
; OTHER INFORMATION: Ceres Seq. ID no. 14302365
US-11-096-568A-12579

Query Match 73.8%; Score 31; DB 7; Length 76;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AEGSPGL 8
Db 43 ASGSPGL 49

RESULT 71
US-11-087-099-8268
; Sequence 8268, Application US/11087099
```

```
; Publication No. US20060041961a1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450)B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 8268
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-11-087-099-8268

Query Match      73.8%; Score 31; DB 7; Length 219;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AEGSPGL 8
DB      3 ARGSPGL 9

RESULT 72
US-11-072-512-3028
; Sequence 3028, Application US/11072512
; Publication No. US20060029945A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: Novel full length cDNA
; CURRENT APPLICATION NUMBER: US/11/072,512
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/350,978
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: JP 2001-379298
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3028
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-072-512-3028

Query Match      73.8%; Score 31; DB 7; Length 285;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AEGSPGL 8
DB      39 ASGSPGL 45

RESULT 73
US-10-858-730-83

; Sequence 83, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Thermobifida fusca
US-10-858-730-83

Query Match      73.8%; Score 31; DB 6; Length 388;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 AEGSPGL 8
DB      351 AEGAPGI 357

RESULT 74
US-11-115-868-2
; Sequence 2, Application US/11115868
; Publication No. US20050266554A1
; GENERAL INFORMATION:
; APPLICANT: D'Amour, Kevin Allen
; APPLICANT: Agulnick, Alan D.
; APPLICANT: Eliazar, Susan
; APPLICANT: Baetge, Emmanuel E.
; TITLE OF INVENTION: PDX1 EXPRESSING ENDODERM
; FILE REFERENCE: CYTHERA.043A
; CURRENT APPLICATION NUMBER: US/11/115,868
; CURRENT FILING DATE: 2005-04-26
; PRIOR APPLICATION NUMBER: 11/021618
; PRIOR FILING DATE: 2004-12-23
; PRIOR APPLICATION NUMBER: 60/587942
; PRIOR FILING DATE: 2004-07-14
; PRIOR APPLICATION NUMBER: 60/586566
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/566293
; PRIOR FILING DATE: 2004-04-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-115-868-2

Query Match      73.8%; Score 31; DB 7; Length 414;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GAEGSPG 7
```

Db 298 GAGSPG 304
||| |||

RESULT 75
US-11-165-305-2
; Sequence 2, Application US/11165305
; Publication No. US20060003313A1
; GENERAL INFORMATION:
; APPLICANT: Kevin Allen D'Amour
; APPLICANT: Alan D. Agulnick
; APPLICANT: Susan Eliazar
; APPLICANT: Emmanuel E. Baetge
; TITLE OF INVENTION: METHODS FOR IDENTIFYING FACTORS FOR
; TITLE OF INVENTION: DIFFERENTIATING DEFINITIVE ENCODER
; FILE REFERENCE: CYTHERA.045CP1
; CURRENT APPLICATION NUMBER: US/11/165.305
; CURRENT FILING DATE: 2005-06-23
; PRIOR APPLICATION NUMBER: 60/532004
; PRIOR FILING DATE: 2003-12-23
; PRIOR APPLICATION NUMBER: 60/566293
; PRIOR FILING DATE: 2004-04-27
; PRIOR APPLICATION NUMBER: 60/586566
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/587942
; PRIOR FILING DATE: 2004-07-14
; PRIOR APPLICATION NUMBER: 11/021618
; PRIOR FILING DATE: 2004-12-23
; PRIOR APPLICATION NUMBER: 11/115868
; PRIOR FILING DATE: 2005-04-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-165-305-2

Query Match 73.8%; Score 31; DB 7; Length 414;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GAGSPG 7
||| |||
Db 298 GAGSPG 304

Search completed: March 11, 2006, 12:12:00
Job time : 16.2 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 11:51:11 ; Search time 319.8 Seconds
(without alignments)
16.487 Million cell updates/sec

Title: US-10-698-121A-2
Perfect score: 62
Sequence: 1 GEKGAEGSPGLL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : A Geneseq 21.*

- 1: geneseqp1980s.*
- 2: geneseqp1990s.*
- 3: geneseqp2000s.*
- 4: geneseqp2001s.*
- 5: geneseqp2002s.*
- 6: geneseqp2003as.*
- 7: geneseqp2003bs.*
- 8: geneseqp2004s.*
- 9: geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	12	AD015279	Ad015279 Collagen
2	58	93.5	652	ABM83520	Abm83520 Human dia
3	58	93.5	654	ABM83519	Abm83519 Human dia
4	58	93.5	657	ABM83518	Abm83518 Human dia
5	58	93.5	661	ABM83517	Abm83517 Human dia
6	58	93.5	673	ABM83516	Abm83516 Human dia
7	58	93.5	683	ABM83515	Abm83515 Human dia
8	58	93.5	685	ABM83514	Abm83514 Human dia
9	58	93.5	688	ABM83529	Abm83529 Human dia
10	58	93.5	692	ABM83528	Abm83528 Human dia
11	58	93.5	695	ABM83527	Abm83527 Human dia
12	58	93.5	697	ABM83526	Abm83526 Human dia
13	58	93.5	699	ABM83525	Abm83525 Human dia
14	58	93.5	700	ABM83524	Abm83524 Human dia
15	58	93.5	702	ABM83522	Abm83522 Human dia
16	58	93.5	702	ABM83523	Abm83523 Human dia
17	58	93.5	705	ABR40108	Abm83521 Human dia
18	58	93.5	714	ABM83521	Abm83521 Human dia
19	58	93.5	717	ADN95515	Adn95515 Human BEC
20	52	83.9	93	AAB34143	Aab34143 Gene 1 hu
21	52	83.9	1603	ADQ21405	Adq21405 Human sof
22	51	82.3	352	ADY65014	Ady65014 S. manson
23	50	80.6	675	ABM82963	Abm82963 Human dia
24	50	80.6	699	ABM82962	Abm82962 Human dia

25	50	80.6	703	8	ABM82961	Abm82961 Human dia
26	50	80.6	705	2	AAV08305	AAV08305 Human col
27	49	79.0	232	9	ADY66076	Ady66076 S. manson
28	49	79.0	244	8	ADF72542	Adf72542 Human APM
29	49	79.0	311	4	AAW41798	AAW41798 Human pol
30	49	79.0	549	2	AAW40113	AAW40113 Human alp
31	49	79.0	994	8	ABM82875	Abm82875 Human dia
32	49	79.0	1061	8	ABM82874	Abm82874 Human dia
33	49	79.0	1078	4	AAW40012	AAW40012 Human pol
34	49	79.0	1712	5	AAO17361	AAO17361 Human alp
35	49	79.0	1712	8	ADS10473	ADS10473 Human the
36	49	79.0	1712	8	ADU06526	ADU06526 Novel bro
37	49	79.0	1712	9	ADY59885	Ady59885 Human col
38	48	77.4	293	6	ABU70813	Abu70813 Human adi
39	48	77.4	309	4	AAU23675	AAU23675 Novel hum
40	48	77.4	309	4	ABG60248	ABG60248 Human ova
41	48	77.4	309	5	ABG61719	ABG61719 Novel ova
42	48	77.4	339	6	ABU70735	Abu70735 Human adi
43	48	77.4	623	2	AAW12843	AAW12843 Pro-alpha
44	48	77.4	626	2	AAW12842	AAW12842 Truncated
45	48	77.4	1028	8	ADN35278	Adn35278 Helical d
46	48	77.4	1078	2	AAW1704	AAW1704 Collagen
47	48	77.4	1078	3	AAV96125	AAV96125 Collagen
48	48	77.4	1078	5	AAE16478	AAE16478 Human col
49	48	77.4	1078	5	ABB80736	ABB80736 Collagen
50	48	77.4	1078	5	ABB09628	ABB09628 Amino aci
51	48	77.4	1078	7	ADF13078	Adf13078 Human col
52	48	77.4	1196	2	AAW28916	AAW28916 Type III
53	48	77.4	1313	8	ADN35279	Adn35279 Synthetic
54	48	77.4	1313	8	ADN35277	Adn35277 Helical d
55	48	77.4	1466	4	AAE02537	AAE02537 Porcine a
56	48	77.4	1466	4	AAE02534	AAE02534 Bovine al
57	48	77.4	1466	4	AAE02533	AAE02533 Bovine al
58	48	77.4	1466	4	ABBS0291	ABBS0291 Collagen
59	48	77.4	1466	5	ABBS0747	ABBS0747 Human Tum
60	48	77.4	1466	6	ABU54454	ABU54454 Human Tum
61	48	77.4	1466	6	ABR47418	ABR47418 Breast ca
62	48	77.4	1466	7	ADP65248	ADP65248 Human alp
63	48	77.4	1466	7	ADP65210	ADP65210 Human alp
64	48	77.4	1466	8	ADQ26091	Adq26091 Type III
65	48	77.4	1466	8	ADQ29677	Adq29677 Human col
66	48	77.4	1466	8	ADRI6802	Adri6802 Human col
67	48	77.4	1466	8	ADRI6427	Adri6427 Human col
68	48	77.4	1466	8	ABM80366	Abm80366 Tumour-as
69	48	77.4	1466	8	ADR67267	Adr67267 Human bla
70	48	77.4	1466	9	ADV70235	Adv70235 Tumox-ass
71	48	77.4	1466	9	ADZ09873	Adz09873 Human bre
72	48	77.4	1466	9	AEA04495	Aea04495 Human pro
73	48	77.4	1466	9	ABG15191	Abg15191 Novel hum
74	48	77.4	1470	7	ADU09399	Adu09399 Novel pro
75	48	77.4	1572	8	ADU04510	Adu04510 SLPI-Coll
76	48	77.4	1604	8	ADU04497	Adu04497 Modified
77	48	77.4	1726	6	ABR42661	AbR42661 Decorin-m
78	48	77.4	1950	8	ADU04493	Adu04493 Modified
79	47	75.8	153	9	ADY78066	Ady78066 Human adi
80	47	75.8	166	9	ADY78067	Ady78067 Human adi
81	47	75.8	187	7	ADY78067	Ady78067 Human adi
82	47	75.8	187	7	ADY78067	Ady78067 Human adi
83	47	75.8	187	7	ADY78067	Ady78067 Human adi
84	47	75.8	187	7	ADY78067	Ady78067 Human adi
85	47	75.8	187	7	ADY78067	Ady78067 Human adi
86	47	75.8	187	7	ADY78067	Ady78067 Human adi
87	47	75.8	187	7	ADY78067	Ady78067 Human adi
88	47	75.8	187	7	ADY78067	Ady78067 Human adi
89	47	75.8	187	7	ADY78067	Ady78067 Human adi
90	47	75.8	187	7	ADY78067	Ady78067 Human adi
91	47	75.8	187	7	ADY78067	Ady78067 Human adi
92	47	75.8	187	7	ADY78067	Ady78067 Human adi
93	47	75.8	187	7	ADY78067	Ady78067 Human adi
94	47	75.8	187	7	ADY78067	Ady78067 Human adi
95	47	75.8	187	7	ADY78067	Ady78067 Human adi
96	47	75.8	187	7	ADY78067	Ady78067 Human adi
97	47	75.8	187	7	ADY78067	Ady78067 Human adi

98	47	75.8	203	7	ADE93622	Adc93622 Human adi
99	47	75.8	203	7	ADE93640	Adc93640 Human adi
100	47	75.8	203	7	ADE93642	Adc93642 Human adi
101	47	75.8	203	7	ADE93605	Adc93605 Human adi
102	47	75.8	226	7	ADE93604	Adc93604 Human adi
103	47	75.8	231	2	AAY21808	Aay21808 Adipose m
104	47	75.8	231	4	AAB60347	Aab60347 Recombina
105	47	75.8	233	7	ADD93530	Add93530 Novel NOV
106	47	75.8	243	7	ADP28492	Adp28492 Adiponect
107	47	75.8	243	8	ADP88008	Adp88008 Monkey ad
108	47	75.8	243	8	ADU71354	Adu71354 Rhesus mo
109	47	75.8	244	2	AAW09108	Aaw09108 Human adi
110	47	75.8	244	2	AAW21807	Aay21807 Adipose m
111	47	75.8	244	3	AAV71053	Aay71053 Human APM
112	47	75.8	244	3	AAV71035	Aay71035 Human APM
113	47	75.8	244	3	AAV85144	Aay85144 Human adi
114	47	75.8	244	3	AAB30233	Aab30233 Human adi
115	47	75.8	244	4	AAB65828	Aab65828 Human adi
116	47	75.8	244	4	AAE05529	Aae05529 Human OBG
117	47	75.8	244	4	AAB49598	Aab49598 Human ACR
118	47	75.8	244	4	AAB50373	Aab50373 Human adi
119	47	75.8	244	4	AAB49592	Aab49592 Human ACR
120	47	75.8	244	5	ABB08223	Abb08223 Human apm
121	47	75.8	244	5	AAG80254	Aag80254 Human APM
122	47	75.8	244	6	ABP70915	Abp70915 APM1 prot
123	47	75.8	244	6	ABR42035	Abc42035 Human APM
124	47	75.8	244	6	ABU98378	Abu98378 Human adi
125	47	75.8	244	6	ABP70905	Abp70905 APM1 prot
126	47	75.8	244	6	AAO16570	Aao16570 APM1 secr
127	47	75.8	244	6	ABR39800	Abc39800 APM1 poly
128	47	75.8	244	6	ABR40088	Abc40088 APM1 prot
129	47	75.8	244	6	ABR40249	Abc40249 Human APM
130	47	75.8	244	6	AAO30071	Aao30071 Human OBG
131	47	75.8	244	6	AAE37142	Aae37142 Human APM
132	47	75.8	244	6	ABR40221	Abc40221 Human gen
133	47	75.8	244	6	ABR39850	Abc39850 APM1 poly
134	47	75.8	244	6	ABR44306	Abc44306 APM1 poly
135	47	75.8	244	6	ABP70910	Abp70910 APM1 prot
136	47	75.8	244	6	ABP79854	Abp79854 Amino aci
137	47	75.8	244	6	ABR40011	Abc40011 Human APM
138	47	75.8	244	6	ABR43906	Abc43906 APM1 poly
139	47	75.8	244	6	ABU08790	Abu08790 Adiponect
140	47	75.8	244	6	ABP98485	Abp98485 Amino aci
141	47	75.8	244	6	ADA00739	Ada00739 APM1 poly
142	47	75.8	244	6	ABR40142	Abc40142 APM1 prot
143	47	75.8	244	6	ABR40083	Abc40083 APM1 prot
144	47	75.8	244	6	ABR40216	Abc40216 Human cob
145	47	75.8	244	6	ABR39864	Abc39864 APM1 poly
146	47	75.8	244	6	ABR40296	Abc40296 Human APM
147	47	75.8	244	6	AAO16579	Aao16579 Energen-r
148	47	75.8	244	6	ABG79989	Abg79989 Human APM
149	47	75.8	244	6	ABP71421	Abp71421 APM1 poly
150	47	75.8	244	6	ABR63590	Abc63590 APM1 prot
151	47	75.8	244	6	ABR61851	Abc61851 REDAX pol
152	47	75.8	244	6	ABR82228	Abc82228 Glucoset
153	47	75.8	244	6	ABR61768	Abc61768 PROLIXIN
154	47	75.8	244	6	AAO29685	Aao29685 Human APM
155	47	75.8	244	6	AAO29601	Aao29601 Human APM
156	47	75.8	244	6	AAO29593	Aao29593 Human APM
157	47	75.8	244	7	ABU62127	Abu62127 Human adi
158	47	75.8	244	7	ABR63585	Abc63585 APM1 prot
159	47	75.8	244	7	ABR61798	Abc61798 RYZN poly
160	47	75.8	244	7	ABR82288	Abc82288 BROMIX li
161	47	75.8	244	7	ABR61980	Abc61980 DEXAR pol
162	47	75.8	244	7	ADC01979	Adc01979 Human adi
163	47	75.8	244	7	ADC21967	Adc21967 Human ins
164	47	75.8	244	7	ADD93522	Add93522 Novel NOV
165	47	75.8	244	7	ADD93526	Add93526 Novel NOV
166	47	75.8	244	7	ADD93524	Add93524 Novel NOV
167	47	75.8	244	7	ADF16684	Adf16684 Human alb
168	47	75.8	244	7	ADF16683	Adf16683 Human alb
169	47	75.8	244	7	ADP16685	Adp16685 Human alb
170	47	75.8	244	7	ADP28494	Adp28494 Human adi

171	47	75.8	244	7	ADE93603	Adc93603 Human adi
172	47	75.8	244	7	ADF94864	Adf94864 Human gen
173	47	75.8	244	7	ADG89987	Adg89987 Human ACR
174	47	75.8	244	7	ADH21884	Adh21884 Human adi
175	47	75.8	244	7	ADH21886	Adh21886 Human HDA
176	47	75.8	244	7	ADH21885	Adh21885 Human HDA
177	47	75.8	244	7	ABU64235	Abu64235 APM1 - li
178	47	75.8	244	7	ADM66423	Adm66423 Human adi
179	47	75.8	244	8	ADF72538	Adf72538 Human OBG
180	47	75.8	244	8	ADF72549	Adf72549 Human APM
181	47	75.8	244	8	ADG14699	Adg14699 Human OBG
182	47	75.8	244	8	ADF14854	Adf14854 Human art
183	47	75.8	244	8	ADL26858	Adl26858 Human adi
184	47	75.8	244	8	ADO47938	Ado47938 Human adi
185	47	75.8	244	8	ADQ18231	Adq18231 Human sof
186	47	75.8	244	8	ADP88006	Adp88006 Human adi
187	47	75.8	244	8	ADS20396	Ads20396 Human adi
188	47	75.8	244	8	ADU23372	Adu23372 Human Adi
189	47	75.8	244	8	ADU51248	Adu51248 Human adi
190	47	75.8	244	8	ADU71352	Adu71352 Human adi
191	47	75.8	244	8	ADV94918	Adv94918 Human fat
192	47	75.8	244	9	ADV14625	Adv14625 PPARGamma
193	47	75.8	244	9	ADY20832	Ady20832 Human ACR
194	47	75.8	244	9	ADY78064	Ady78064 Human adi
195	47	75.8	244	9	AEA46586	Aea46586 Human ACR
196	47	75.8	250	7	ADD93528	Add93528 Novel NOV
197	47	75.8	685	8	ADU02276	Adu02276 Novel hum
198	47	75.8	772	2	AAR23873	Aar23873 Human alp
199	47	75.8	772	2	AAW09643	Aaw09643 Human typ
200	47	75.8	836	7	ADF16522	Adf16522 Human alb
201	47	75.8	836	7	ADH21811	Adh21811 Human alb
202	47	75.8	1152	8	ADQ59439	Adq59439 Human can
203	47	75.8	1152	9	ADZ13746	Adz13746 Murine ca
204	47	75.8	1268	8	ABO84442	Ab084442 Human can
205	47	75.8	1283	8	ABO84443	Ab084443 Human can
206	47	75.8	1685	4	ABG04839	Abg04839 Novel hum
207	47	75.8	1685	8	ABO84444	Ab084444 Human can
208	47	75.8	1693	4	ABG15619	Abg15619 Novel hum
209	46	74.2	182	3	AB32814	Ab32814 Eucalyptu
210	46	74.2	194	8	ADP88009	Adp88009 Dog adipo
211	46	74.2	441	3	AAV50914	Aay50914 Human fet
212	46	74.2	518	4	ABG22048	Abg22048 Novel hum
213	46	74.2	635	9	ADM17297	Adm17297 Eucalyptu
214	46	74.2	635	9	ADM17298	Adm17298 Eucalyptu
215	46	74.2	714	6	ABE36862	Abc36862 Human col
216	46	74.2	714	8	ADI33185	Adi33185 Human col
217	46	74.2	926	7	ABO83805	Ab083805 Pseudomon
218	46	74.2	1014	7	ADC21591	Adc21591 Mouse typ
219	46	74.2	1014	7	ADC21544	Adc21544 Human typ
220	46	74.2	1053	4	ABE85863	Abc85863 Murine ad
221	46	74.2	1056	9	AEC04787	Aec04787 Human bre
222	46	74.2	1081	9	AEC04183	Aec04183 Human bre
223	46	74.2	1417	8	ABM83560	Abm83560 Human dia
224	46	74.2	1418	2	AAR59751	Aar59751 Type II c
225	46	74.2	1418	2	AAR71703	Aar71703 Collagen
226	46	74.2	1418	3	AAV96124	Aay96124 Collagen
227	46	74.2	1418	4	AAV35624	Aay35624 Human typ
228	46	74.2	1418	5	AAE16477	Aae16477 Human col
229	46	74.2	1418	5	ABB80735	Abb80735 Collagen
230	46	74.2	1418	5	ABG93927	Abg93927 Human pol
231	46	74.2	1418	5	ABB09627	Abb09627 Amino aci
232	46	74.2	1418	7	ADF13077	Adf13077 Human col
233	46	74.2	1418	9	ABE25695	Abc25695 Human col
234	46	74.2	1419	7	ADE55694	Adc55694 Rat Prote
235	46	74.2	1420	8	ADP48530	Adp48530 Full-leng
236	46	74.2	1442	2	AAR79480	Aar79480 Rat type
237	46	74.2	1487	2	AAW61562	Aaw61562 Human typ
238	46	74.2	1487	5	ABG61861	Abg61861 Prostatac
239	46	74.2	1487	5	ABG93928	Abg93928 Mouse pol
240	46	74.2	1487	6	ABP56769	Abp56769 Collagen
241	46	74.2	1487	9	ADX15809	Adx15809 Human Col
242	46	74.2	1629	7	ADP65203	Adp65203 Human alp
243	46	74.2	1669	5	ABB57334	Abb57334 Mouse isc

244	46	74.2	1691	9	ADW44472	Adw44472 Murine pr	317	45	72.6	516	7	ADe87046	Human pan
245	46	74.2	1739	4	AG77792	Aeq77792 Murine pr	318	45	72.6	544	8	ADm48393	Recombina
246	46	74.2	1744	8	ADG23179	Adg23179 Bacterial	319	45	72.6	595	2	AAy06239	Mouse rec
247	46	74.2	1745	5	ADG77793	Aeg77793 Human pro	320	45	72.6	660	9	ADW99570	Human gel
248	46	74.2	1745	5	ABW97234	Abw97234 Novel hum	321	45	72.6	660	9	ADW99571	Human gel
249	46	74.2	1745	8	ADQ19841	Adq19841 Human sof	322	45	72.6	662	4	ADW68072	Amino aci
250	46	74.2	1767	8	ADQ39813	Adq39813 Human myo	323	45	72.6	662	4	AAE02718	Human alp
251	46	74.2	1767	8	ADQ39817	Adq39817 Human myo	324	45	72.6	662	4	ADW68072	Human alp
252	46	74.2	1806	5	AAU84266	Aau84266 Human end	325	45	72.6	723	4	ABW62129	Recombina
253	46	74.2	1806	5	ABW05596	Abw05596 Breast ca	326	45	72.6	822	2	AAy06240	Mouse rec
254	46	74.2	1806	6	ABW58545	Abw58545 Human can	327	45	72.6	886	7	ADW6240	Human pan
255	46	74.2	1806	6	ABW58545	Abw58545 Human can	328	45	72.6	886	7	ADW6240	Human pan
256	46	74.2	1806	6	ABW58545	Abw58545 Human can	329	45	72.6	957	4	ADW52683	Human gel
257	46	74.2	1806	6	ADP65251	Adp65251 Lung canc	330	45	72.6	1014	9	ADW99575	Human gel
258	46	74.2	1806	8	ADQ39815	Adq39815 Human myo	331	45	72.6	1014	9	ADW99577	Human gel
259	46	74.2	1806	8	ADQ39815	Adq39815 Human myo	332	45	72.6	1014	9	ADW99577	Human gel
260	46	74.2	1806	9	ADQ39815	Adq39815 Human myo	333	45	72.6	1014	9	ADW99572	Human gel
261	46	74.2	1806	9	ADQ39815	Adq39815 Human myo	334	45	72.6	1014	9	ADW99576	Human gel
262	46	74.2	1818	8	ADQ39812	Adq39812 Human myo	335	45	72.6	1014	9	ADW99578	Human gel
263	46	74.2	1818	8	ADQ39814	Adq39814 Human myo	336	45	72.6	1014	9	ADW99580	Human gel
264	45	72.6	62	2	AAW05303	Aaw05303 Collagen	337	45	72.6	1057	3	AAy84541	Human pro
265	45	72.6	63	5	AAU75471	Aau75471 Collagen	338	45	72.6	1057	3	AAy84541	Amino aci
266	45	72.6	63	5	ABW43719	Abw43719 Collagen	339	45	72.6	1058	3	AAy84544	A human c
267	45	72.6	63	8	ADJ55711	Adj55711 Collagen	340	45	72.6	1107	2	AAW89472	Collagen/
268	45	72.6	63	8	ADJ55711	Adj55711 Collagen	341	45	72.6	1107	2	AAW89472	Collagen/
269	45	72.6	63	9	AE811736	Aeb11736 Collagen-	342	45	72.6	1107	3	AAy84540	Amino aci
270	45	72.6	63	9	AE847080	Aeb47080 Engineere	343	45	72.6	1161	7	ADW87050	Human pan
271	45	72.6	69	4	AAW95146	Aaw95146 Collagen	344	45	72.6	1169	2	AAW89469	Collagen/
272	45	72.6	69	4	AAW95146	Aaw95146 Collagen	345	45	72.6	1169	2	AAW89469	Collagen/
273	45	72.6	69	4	AAW95146	Aaw95146 Collagen	346	45	72.6	1169	2	AAW89469	Collagen/
274	45	72.6	72	2	AAW95148	Aaw95148 Collagen	347	45	72.6	1171	2	AAW89470	Collagen/
275	45	72.6	72	4	AAW95148	Aaw95148 Collagen	348	45	72.6	1171	2	AAW89470	Collagen/
276	45	72.6	72	4	AAW95148	Aaw95148 Collagen	349	45	72.6	1212	7	ADW89022	Human pan
277	45	72.6	82	2	AAW95149	Aaw95149 Collagen	350	45	72.6	1212	7	ADW89022	Human pan
278	45	72.6	85	4	AAW95149	Aaw95149 Collagen	351	45	72.6	1212	7	ADW89022	Human pan
279	45	72.6	85	4	AAW95149	Aaw95149 Collagen	352	45	72.6	1212	7	ADW89022	Human pan
280	45	72.6	85	4	AAW95149	Aaw95149 Collagen	353	45	72.6	1212	7	ADW89022	Human pan
281	45	72.6	112	2	AAW95149	Aaw95149 Collagen	354	45	72.6	1212	7	ADW89022	Human pan
282	45	72.6	167	4	AAW95149	Aaw95149 Collagen	355	45	72.6	1212	7	ADW89022	Human pan
283	45	72.6	167	4	AAW95149	Aaw95149 Collagen	356	45	72.6	1212	7	ADW89022	Human pan
284	45	72.6	250	8	ADW91505	Adw91505 Partial h	357	45	72.6	1212	7	ADW89022	Human pan
285	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	358	45	72.6	1212	7	ADW89022	Human pan
286	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	359	45	72.6	1212	7	ADW89022	Human pan
287	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	360	45	72.6	1212	7	ADW89022	Human pan
288	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	361	45	72.6	1212	7	ADW89022	Human pan
289	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	362	45	72.6	1212	7	ADW89022	Human pan
290	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	363	45	72.6	1212	7	ADW89022	Human pan
291	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	364	45	72.6	1212	7	ADW89022	Human pan
292	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	365	45	72.6	1212	7	ADW89022	Human pan
293	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	366	45	72.6	1212	7	ADW89022	Human pan
294	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	367	45	72.6	1212	7	ADW89022	Human pan
295	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	368	45	72.6	1212	7	ADW89022	Human pan
296	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	369	45	72.6	1212	7	ADW89022	Human pan
297	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	370	45	72.6	1212	7	ADW89022	Human pan
298	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	371	45	72.6	1212	7	ADW89022	Human pan
299	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	372	45	72.6	1212	7	ADW89022	Human pan
300	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	373	45	72.6	1212	7	ADW89022	Human pan
301	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	374	45	72.6	1212	7	ADW89022	Human pan
302	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	375	45	72.6	1212	7	ADW89022	Human pan
303	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	376	45	72.6	1212	7	ADW89022	Human pan
304	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	377	45	72.6	1212	7	ADW89022	Human pan
305	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	378	45	72.6	1212	7	ADW89022	Human pan
306	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	379	45	72.6	1212	7	ADW89022	Human pan
307	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	380	45	72.6	1212	7	ADW89022	Human pan
308	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	381	45	72.6	1212	7	ADW89022	Human pan
309	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	382	45	72.6	1212	7	ADW89022	Human pan
310	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	383	45	72.6	1212	7	ADW89022	Human pan
311	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	384	45	72.6	1212	7	ADW89022	Human pan
312	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	385	45	72.6	1212	7	ADW89022	Human pan
313	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	386	45	72.6	1212	7	ADW89022	Human pan
314	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	387	45	72.6	1212	7	ADW89022	Human pan
315	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	388	45	72.6	1212	7	ADW89022	Human pan
316	45	72.6	251	4	AAW68071	Aaw68071 Amino aci	389	45	72.6	1212	7	ADW89022	Human pan

330	45	72.6	1464	9	ADV87000	Collagen	463	43	69.4	423	2	AAR99306	Aar99306	Bluegill
331	45	72.6	1464	9	ADV70233	Tumor-ass	464	43	69.4	433	8	ABM82895	ABM82895	Human dia
332	45	72.6	1464	9	AD58061	Amno aci	465	43	69.4	469	7	ADE61178	AdE61178	Rat Prote
333	45	72.6	1464	9	AD225573	Human typ	466	43	69.4	482	9	ADV86420	Adv86420	Novel Clq
334	45	72.6	1464	9	ADZ09874	Human bre	467	43	69.4	485	8	ADG98282	AdG98282	Mouse Clq
335	45	72.6	1464	9	AEA04480	Human pro	468	43	69.4	494	9	ADV86416	Adv86416	Novel Clq
336	45	72.6	1536	2	AD87051	Human pan	469	43	69.4	495	2	AAW39747	Aaw39747	Human mar
337	45	72.6	1574	2	AY30680	Splice va	470	43	69.4	520	2	AAW39748	Aaw39748	Human mar
338	45	72.6	1669	4	AM40863	Human pol	471	43	69.4	520	2	AAW39748	Aaw39748	Human mar
339	45	72.6	1669	5	ABB90760	Human tum	472	43	69.4	520	3	AAW39748	Aaw39748	Human mar
400	45	72.6	1669	6	ABU54467	Human tum	473	43	69.4	520	3	AAW39748	Aaw39748	Human mar
401	45	72.6	1669	7	ADF90900	Human hep	474	43	69.4	520	5	ABH81159	ABH81159	Human WAR
402	45	72.6	1669	8	ADR87604	Human Typ	475	43	69.4	520	6	ABU58565	ABU58565	Human PRO
403	45	72.6	1672	4	AM39077	Human pol	476	43	69.4	520	6	ABU88113	ABU88113	Novel hum
404	45	72.6	1779	4	ABB60207	Drosophil	477	43	69.4	520	6	ABU84428	ABU84428	Human sec
405	44	71.0	139	9	ADY64898	S. manson	478	43	69.4	520	6	ABR66302	ABr66302	Human sec
406	44	71.0	229	4	ABB69906	Drosophil	479	43	69.4	520	6	ABR65692	ABr65692	Human sec
407	44	71.0	265	9	AD222859	INSP16-B	480	43	69.4	520	6	ABU99632	ABU99632	Human sec
408	44	71.0	271	9	AD222869	His-tagge	481	43	69.4	520	6	ABU82871	ABU82871	Human PRO
409	44	71.0	278	9	ADY65012	S. manson	482	43	69.4	520	6	ABU89992	ABU89992	Novel hum
410	44	71.0	294	9	ADZ22855	INSP161 m	483	43	69.4	520	6	ABR68241	ABr68241	Human sec
411	44	71.0	300	9	ADZ22865	INSP161 H	484	43	69.4	520	6	ABU96294	ABU96294	Novel hum
412	44	71.0	400	8	ADP04725	Sea squir	485	43	69.4	520	6	ABU92725	ABU92725	Human sec
413	44	71.0	410	9	ADZ22857	INSP16-A	486	43	69.4	520	6	ABO08802	ABO08802	Human sec
414	44	71.0	416	9	ADZ22867	His-tagge	487	43	69.4	520	6	ABO02854	ABO02854	Human sec
415	44	71.0	438	9	ADV86405	Novel Clq	488	43	69.4	520	6	ABR75008	ABr75008	Human sec
416	44	71.0	459	7	ADH89023	TAR263.1	489	43	69.4	520	6	ABR94770	ABr94770	Human sec
417	44	71.0	459	7	ADJ70748	Human hea	490	43	69.4	520	6	ABO25278	ABO25278	Novel hum
418	44	71.0	459	8	ADR87626	Human Typ	491	43	69.4	520	6	ABU85743	ABU85743	Human PRO
419	44	71.0	459	9	ADY86382	Novel Clq	492	43	69.4	520	6	ABU89903	ABU89903	Novel hum
420	44	71.0	459	9	ADY70232	Tumor-ass	493	43	69.4	520	6	ABU98118	ABU98118	Novel hum
421	44	71.0	476	9	ADZ22875	INSP161 s	494	43	69.4	520	6	ABU91824	ABU91824	Novel hum
422	44	71.0	476	9	ADZ22877	His-tagge	495	43	69.4	520	6	ABU72284	ABU72284	Novel hum
423	44	71.0	477	9	ADY86418	Novel Clq	496	43	69.4	520	6	ABU89517	ABU89517	Human PRO
424	44	71.0	553	8	ADF69307	Human lun	497	43	69.4	520	6	ABU86358	ABU86358	Human sec
425	44	71.0	567	8	ADT51054	Cancer re	498	43	69.4	520	6	ABU67571	ABU67571	Human PRO
426	44	71.0	616	4	ASG19456	Novel hum	499	43	69.4	520	6	ABU80599	ABU80599	Human PRO
427	44	71.0	640	8	ADT51053	Cancer re	500	43	69.4	520	6	ABR99517	ABR99517	Human sec
428	44	71.0	673	8	ADJ11751	Human the								
429	44	71.0	680	5	ABJ05575	Breast ca								
430	44	71.0	680	6	ABU56618	Lung canc								
431	44	71.0	680	6	ABR47414	Breast ca								
432	44	71.0	680	7	ADN46128	Human Pro								
433	44	71.0	680	7	ADN38792	Cancer/an								
434	44	71.0	680	8	ADL06538	Human tum								
435	44	71.0	680	8	ADV15274	Human ost								
436	44	71.0	680	8	ADV15279	Human ost								
437	44	71.0	680	9	ADZ09776	Human bre								
438	44	71.0	680	9	AEA00131	Human TAT								
439	44	71.0	680	9	AEA00651	Human TAT								
440	44	71.0	700	5	ADP69082	Human pol								
441	44	71.0	702	8	ADJ67574	Human ova								
442	44	71.0	738	9	ADW44464	Murine co								
443	44	71.0	842	8	ABO84441	Mouse can								
444	44	71.0	900	5	ABP43610	Alpha-2 t								
445	44	71.0	1051	4	ABG20128	Novel hum								
446	44	71.0	1372	7	ADZ59683	Rat Prote								
447	44	71.0	1372	7	ADN45148	Rat Prote								
448	44	71.0	1372	7	ADN45604	Rat Prote								
449	44	71.0	1372	7	ADZ59687	Rat Prote								
450	44	71.0	1372	7	ADN47529	Rat Prote								
451	44	71.0	1669	9	ADW44466	Murine pr								
452	43	69.4	164	8	ADS98147	Protein f								
453	43	69.4	200	3	AAW43239	Human ORF								
454	43	69.4	206	8	ADS98557	Protein f								
455	43	69.4	210	8	ADS98079	Protein f								
456	43	69.4	216	6	ABR58603	Human can								
457	43	69.4	234	7	ADU62236	Collagen-								
458	43	69.4	285	9	ABC01333	Bovine hC								
459	43	69.4	335	8	ADL91149	Human fic								
460	43	69.4	335	8	ADT77838	Ficolin 1								
461	43	69.4	403	3	AAW58351	Lung canc								
462	43	69.4	407	8	ADQ39967	Human myo								

ALIGNMENTS

RESULT 1

AD015279
ID AD015279 standard; peptide; 12 AA.

XX AC AD015279;

XX DT 12-AUG-2004 (first entry)

XX DE Collagen XIII and alphabeta1 integrin binding inhibiting peptide SEQ.2.

XX KW chronic inflammatory disease; blocking agent;
KW Collagen XIII and alphabeta1 integrin binding inhibitor; collagen XIII;
KW alphabeta1 integrin; inflammatory disease; antiinflammatory;
KW antipeptidic; antiarthritic; gene therapy; renal fibrosis;
KW lung fibrosis; liver fibrosis; rheumatoid arthritis; psoriasis;
XX experimental colitis; crescentic glomerulonephritis.

OS Synthetic.

XX WO2004041846-A2.

XX 21-MAY-2004.

XX 31-OCT-2003; 2003WO-US034818.

XX 01-NOV-2002; 2002US-0423297P.

XX (BOYS-) BOYS TOWN NAT RES HOSPITAL.

XX Cosgrove D;

PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42171.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 654 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 654;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGAGGSPGL 11
| | | | | | | | | | | |
Db 290 GEGAGGSPGL 300
RESULT 4
ABM83518
ID ABM83518 standard; protein; 657 AA.
XX
AC ABM83518;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3767.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
XX Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI

PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42170.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 657 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 657;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGAGGSPGL 11
| | | | | | | | | | | |
Db 293 GEGAGGSPGL 303
RESULT 5
ABM83517
ID ABM83517 standard; protein; 661 AA.
XX
AC ABM83517;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3766.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
XX Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI

PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42169.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
PS
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorders, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 661 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 661;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEKGAEGSPGL 11
|||||
DB 297 GEKGAEGSPGL 307
RESULT 6
ABM83516
ID ABM83516 standard; protein; 673 AA.
XX
AC ABM83516;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3765.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI

PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42169.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
PS
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 673 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 673;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEKGAEGSPGL 11
|||||
DB 309 GEKGAEGSPGL 319
RESULT 7
ABM83515
ID ABM83515 standard; protein; 683 AA.
XX
AC ABM83515;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3764.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI

```

PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42167.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 683 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 683;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEGGAGGSPGL 11
DB 331 GEGGAGGSPGL 341
RESULT 8
ABM83514
ID ABM83514 standard; protein; 685 AA.
XX
AC ABM83514;
XX
XX 18-NOV-2004 (first entry)
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3763.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
XX
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42166.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page; 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 685 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 685;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEGGAGGSPGL 11
DB 321 GEGGAGGSPGL 331
RESULT 9
ABM83529
ID ABM83529 standard; protein; 688 AA.
XX
AC ABM83529;
XX
XX 18-NOV-2004 (first entry)
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3778.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
XX
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthorne TA, Suchorski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
```

```
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42181.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page: 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 688 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 688;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEKGAGGSPGL 11
DB 350 GEKGAGGSPGL 360
RESULT 10
ABM83528
ID ABM83528 standard; protein; 692 AA.
XX
AC ABM83528;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3777.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
XX
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
DR N-PSDB; ACN42180.
XX
XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
XX Claim 27; Page: 190pp; English.
XX
XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
XX Sequence 692 AA;
SQ
Query Match 93.5%; Score 58; DB 8; Length 692;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEKGAGGSPGL 11
DB 328 GEKGAGGSPGL 338
RESULT 11
ABM83527
ID ABM83527 standard; protein; 695 AA.
XX
AC ABM83527;
XX
XX 18-NOV-2004 (first entry)
XX
XX Human diagnostic and therapeutic pprotein SEQ ID NO:3776.
XX
XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
XX Homo sapiens.
XX
XX WO2004023973-A2.
XX
XX 25-MAR-2004.
XX
XX 12-SEP-2003; 2003WO-US028227.
XX
XX 12-SEP-2002; 2002US-0410259P.
XX
XX 12-SEP-2002; 2002US-0410260P.
XX
XX (INCY-) INCYTE CORP.
XX
XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
XX WPI; 2004-329368/30.
```


PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.

PS Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
 XX selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The present sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm

XX Sequence 699 AA;

Query Match 93.5%; Score 58; DB 8; Length 699;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGKAGGSPGL 11

DB 350 GEGKAGGSPGL 360

RESULT 14

ABM83524

ID ABM83524 standard; protein; 700 AA.

XX AC ABM83524;

XX 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:3773.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.

DR N-PSDB; ACN42176.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful

PT

PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.

PS Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
 XX selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC polynucleotide and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The present sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm

XX Sequence 700 AA;

Query Match 93.5%; Score 58; DB 8; Length 700;
 Best Local Similarity 100.0%; Pred. No. 0.69;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGKAGGSPGL 11

DB 350 GEGKAGGSPGL 360

RESULT 15

ABM83522

ID ABM83522 standard; protein; 702 AA.

XX AC ABM83522;

XX 18-NOV-2004 (first entry)

XX Human diagnostic and therapeutic pprotein SEQ ID NO:3771.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.

DR N-PSDB; ACN42174.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human

PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 XX in gene mapping.
 PS Claim 27; Page; 190pp; English.
 XX
 CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 702 AA;

Query Match 93.5%; Score 58; DB 8; Length 702;
 Best Local Similarity 100.0%; Pred. No. 0.7;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGL 11
 |||||
 Db 350 GKGAGGSPGL 360

RESULT 16
 ABM83523
 ID ABM83523 standard; protein; 702 AA.

AC ABM83523;

XX 18-NOV-2004 (first entry)

DE Human diagnostic and therapeutic pprotein SEQ ID NO:3772.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;

XX WPI; 2004-329368/30.

DR N-ESDB; ACN42175.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 XX in diagnosing a condition, disease or disorder associated with human
 XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or

PT in gene mapping.
 XX Claim 27; Page; 190pp; English.
 XX
 CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 702 AA;

Query Match 93.5%; Score 58; DB 8; Length 702;
 Best Local Similarity 100.0%; Pred. No. 0.7;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGL 11
 |||||
 Db 338 GKGAGGSPGL 348

RESULT 17

ABR40108

ID ABR40108 standard; protein; 705 AA.

XX ABR40108;

XX 04-JUL-2003 (first entry)

XX Human cell adhesion and extracellular matrix protein, CADECM-5.

XX Human, anti-HIV; antiallergic; cerebroprotective; antiparkinsonian;
 KW anticonvulsant; nootropic; neuroprotective; immunosuppressive;
 KW dermatological; anti-inflammatory; cytostatic; antiarteriosclerotic;
 KW gene therapy; cell adhesion; extracellular matrix; CADECM;
 KW immune system disorder; AIDS; allergy; neurological disorder; stroke;
 KW Parkinson's disease; epilepsy; developmental disorder; Down's syndrome;
 KW cerebral palsy; connective tissue disorder; systemic lupus erythematosus;
 KW genetic disorder; Alport's syndrome; cell proliferative disorder; cancer;
 KW atherosclerosis.

XX Homo sapiens.

XX WO2003027230-A2.

XX 03-APR-2003.

XX 02-AUG-2002; 2002WO-US024649.

XX 03-AUG-2001; 2001US-0309964P.

XX 03-AUG-2001; 2001US-0310119P.

XX 17-AUG-2001; 2001US-0313091P.

XX 31-AUG-2001; 2001US-0316771P.

XX 07-SEP-2001; 2001US-0317896P.

XX 21-SEP-2001; 2001US-0324781P.

XX 05-OCT-2001; 2001US-0327606P.

XX 12-OCT-2001; 2001US-0328960P.

XX 09-NOV-2001; 2001US-0344471P.

XX 17-MAY-2002; 2002US-0381291P.

XX (INCY-) INCYTE GENOMICS INC.

PI Burford N, Warren BA, Duggan BM, Mason PM, Richardson TW, Yue H;
 PI Forsythe LJ, Elliott VS, Griffin JA, Gorvad AE, Azimzai Y;
 PI Kallick DA, Xu Y, Honchell CD, Baughn MR, Gietzen KJ, Lee S;
 PI Wallia NK, Tang Y, Nguyen DB, Becha SD, Lee SY, Ramkumar J;
 XX WPI: 2003-354645/33.
 DR N-PSDB; ACC00396.
 XX New human cell adhesion and extracellular matrix proteins (CADECM),
 PT useful for diagnosing, treating or preventing disorders associated with
 PT aberrant CADECM expression, e.g. cancer, AIDS, atherosclerosis, allergies
 PT or stroke.
 XX Claim 1; Page 174-175; 234pp; English.
 XX The present invention relates to novel human cell adhesion and
 CC extracellular matrix proteins, CADECM-1 to CADECM-22, and their coding
 CC sequences (ACC00392-ACC00413 and ABR40104-ABR40125). The coding sequences
 CC and proteins are useful in diagnosing, treating and preventing disorders
 CC associated with aberrant expression of CADECM, such as immune system
 CC disorders (e.g. AIDS or allergies), neurological disorders (e.g. stroke,
 CC Parkinson's disease or epilepsy), developmental disorders (e.g. Down's
 CC syndrome or cerebral palsy), connective tissue disorders (e.g. systemic
 CC lupus erythematosus), genetic disorders (e.g. Alport's syndrome) or cell
 CC proliferative disorders (e.g. cancer or atherosclerosis)
 XX Sequence 705 AA;
 SQ Query Match 93.5%; Score 58; DB 6; Length 705;
 Best Local Similarity 100.0%; Pred. No. 0.7;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEGGAGGSPGL 11
 DB 341 GEGGAGGSPGL 351
 RESULT 18
 ABM83521
 ID ABM83521 standard; protein; 714 AA.
 XX AC ABM83521;
 XX DT 18-NOV-2004 (first entry)
 XX DE Human diagnostic and therapeutic pprotein SEQ ID NO:3770.
 XX KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
 XX OS Homo sapiens.
 XX PN WO2004023973-A2.
 XX PD 25-MAR-2004.
 XX PF 12-SEP-2003; 2003WO-US028227.
 XX PR 12-SEP-2002; 2002US-0410259P.
 XX PR 12-SEP-2002; 2002US-0410260P.
 XX PA (INCY-) INCYTE CORP.
 XX PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Eider LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstein EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kifton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;
 XX WPI: 2004-329368/30.
 DR N-PSDB; ACN42173.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.
 XX Claim 27; Page; 190pp; English.
 XX The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX Sequence 714 AA;
 SQ Query Match 93.5%; Score 58; DB 8; Length 714;
 Best Local Similarity 100.0%; Pred. No. 0.71; Mismatches 0; Indels 0; Gaps 0;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEGGAGGSPGL 11
 DB 350 GEGGAGGSPGL 360
 RESULT 19
 ADN95515
 ID ADN95515 standard; protein; 717 AA.
 XX AC ADN95515;
 XX DT 01-JUL-2004 (first entry)
 XX DE Human BEC/LEC-related protein sequence SeqID438.
 XX KW growth; differentiation; blood endothelial cell; BEC;
 KW lymphatic endothelial cell; LEC; hereditary lymphoedema; VEGFR-3;
 KW lymphatic growth agent; VEGF-C; VEGF-D; antiangiogenic; cytotatic;
 KW vasotropic; antiinflammatory; gene therapy; endothelial cell disorder;
 KW inflammatory disease; cancer metastasis; lymphatic system; human.
 XX OS Homo sapiens.
 XX PN WO2003080640-A1.
 XX PD 02-OCT-2003.
 XX PF 07-MAR-2003; 2003WO-US006900.
 XX PR 07-MAR-2002; 2002US-0363019P.
 XX PA (LUDW-) LUDWIG INST CANCER RES.
 PA (LIGN) LICENTIA LTD.
 XX PI Alitalo K, Makinen T, Petrova T, Saharinen P, Saharinen J;
 XX WPI: 2003-876899/81.
 DR N-PSDB; ADN95516.
 XX Example 1; SEQ ID NO 438; 176pp; English.
 XX This invention relates to a method of differentially modulating the

growth or differentiation of blood endothelial cells (BEC) or lymphatic endothelial cells (LEC) comprises contacting endothelial cells with a composition comprising an agent that differentially modulates blood or lymphatic endothelial cells. Treating hereditary lymphoedema comprises identifying a human subject with lymphoedema and with a mutation in at least one allele of a gene encoding a LEC protein, where the mutation correlates with lymphoedema in human subjects, and with the proviso that the LEC protein is not VEGFR-3; and administering to the subject a composition comprising a lymphatic growth agent selected from VEGF-C or VEGF-D polypeptides and polynucleotides. The invention may be useful for the development of compounds with an antiangiogenic, cytostatic, vasotropic or antiinflammatory activity or for gene therapy. The method is useful in modulating the growth or differentiation of blood endothelial cells or lymphatic endothelial cells, in treating hereditary lymphoedema, in screening for an endothelial cell disorder or predisposition to the disorder or in monitoring the efficacy or toxicity of a drug on endothelial cells. The agent is useful in manufacturing a medicament for the differential modulation of blood vessel endothelial cell or lymphatic vessel endothelial cell growth or differentiation. The lymphatic growth agent may also be used in manufacturing a medicament for the treatment of hereditary lymphoedema resulting from a mutation in a LEC gene or of other diseases involving the lymphatic vessels, such as various inflammatory diseases and cancer metastasis via the lymphatic system. The present sequence is that of a human LEC/BEC differentially expressed protein which is related to the method of the invention. Note: This sequence does not appear in the specification but was obtained by the indexer using the source data given in table 14 of the specification.

Sequence 717 AA;
Query Match 93.5%; Score 58; DB 7; Length 717;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPGL 11
| | | | | | | | | | | |
Db 353 GEGGAEGSPGL 363

RESULT 20
AAB34143
ID AAB34143 standard; protein; 93 AA.
XX
XX AAB34143;
AC
XX
XX 26-JAN-2001 (first entry)
DE
XX
XX Gene 1 human secreted protein homologous amino acid sequence #111.
KW Human; secreted protein; diagnosis; antiarthritic; immunosuppressive;
KW antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; vulnery; gene therapy; autoimmune disease;
KW hyperproliferative disorder; neoplasm; cancer; cardiovascular disorder;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW infection; ocular disorder; wound healing; skin aging; food additive;
KW preservative.
XX
XX Homo sapiens.
OS
XX
XX WO200056755-A1.
FN
XX
XX 28-SEP-2000.
PD
XX
XX 16-MAR-2000; 2000WO-US006830.
PF
XX
XX 19-MAR-1999; 99US-0125361P.
PR
XX
XX 10-DEC-1999; 99US-0169910P.
FR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
PI
XX

DR WPI; 2000-587661/55.
XX
XX New isolated nucleic acid molecules encoding 49 human secreted proteins
PT used for preventing, treating or ameliorating medical conditions, for
PT diagnosing pathological conditions or as food additives or preservatives.
XX
XX Disclosure; Page 383-384; 419pp; English.
XX
XX The polynucleotide sequences given in AAC59449 to AAC59497 encode the
CC human secreted proteins given in AAB34092 to AAB34140. AAB34141 to
CC AAB34216 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissue
CC and cells the genes are expressed in. Examples of activities include:
CC antiarthritic; immunosuppressive; antirheumatic; antiproliferative;
CC cytostatic; cardiac; vasotropic; cerebroprotective; neurotropic;
CC neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
CC and vulnery. The polynucleotides and polypeptides can be used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases, hyperproliferative disorders e.g. neoplasms or
CC cancer of the breast or liver, cardiovascular disorders, cerebrovascular
CC disorders, angiogenesis, nervous system disorders, infections caused by
CC bacteria, viruses and fungi and ocular disorders. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities. AAC59440 to AAC59448 and AAB34091 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 93 AA;
Query Match 83.9%; Score 52; DB 3; Length 93;
Best Local Similarity 81.8%; Pred. No. 0.84;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGGAEGSPGL 11
| | | | | | | | | | | |
Db 78 GEGGAEGSPGV 88
RESULT 21
ADQ21405
ID ADQ21405 standard; protein; 1603 AA.
XX
XX ADQ21405;
AC
XX
XX 26-AUG-2004 (first entry)
DT
XX
XX Human soft tissue sarcoma-upregulated protein - SSQ ID 4225.
DE
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
KW
XX
XX Homo sapiens.
OS
XX
XX WO2004048938-A2.
FN
XX
XX 10-JUN-2004.
PD
XX
XX 26-NOV-2003; 2003WO-US038193.
PF
XX
XX 26-NOV-2002; 2002US-0429739P.
FR
XX
XX (PROT-) PROTEIN DESIGN LABS INC.
PA
XX
XX Aziz N, Ginsburg WM, Zlotnik A;
PI
XX
XX WPI; 2004-441208/41.
DR
XX

PT Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.

XX Example 2; SEQ ID NO 4225; 210pp; English.

XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cyostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC protein of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX Sequence 1603 AA;

Query Match 83.9%; Score 52; DB 8; Length 1603;
 Best Local Similarity 81.8%; Pred. No. 16;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPL 11
 |||||:
 Db 817 GEKGAGSGPGV 827

RESULT 22

ID ADY65014
 ID ADY65014 standard; protein; 352 AA.

XX AC ADY65014;

XX DT 02-JUN-2005 (first entry)

XX DE S. mansoni protein SEQ ID 432.

XX KW Schistosomacide; schistosomiasis; psychiatric disorder; vaccine;
 KW diagnosis.

XX OS Schistosoma mansoni.

XX PN WO2005023979-A2.

XX PD 17-MAR-2005.

XX PF 10-SEP-2004; 2004WO-BR000170.

XX PR 11-SEP-2003; 2003US-0502277P.

XX PA (AMPA-) FUNDACAO AMPARO A PESQUISA DO ESTADO.

XX PI Verjovski-Almeida S, Leite LCC, Farias LP, Miyasato PA, Kawano T;
 PI Demarco R, Garcia JCL, Martins EAL, Ho PL, Nascimento ALTO;
 PI Dias-Neto E, Setubal JC, Menck CFM, Madeira AMBN, Rodrigues V;
 PI Gargioni C;

XX WIPI; 2005-223357/23.

XX New isolated nucleic acid molecule encoding a Schistosoma mansoni
 PT protein, useful for as a vaccine or for preventing, diagnosing, or
 PT treating Schistosoma mansoni infection.

XX Disclosure; SEQ ID NO 432; 52pp; English.

XX The invention relates to an isolated nucleic acid molecule encoding a
 CC Schistosoma mansoni protein, or its portion which is at least 20 amino
 CC acids in length. Also included are an expression vector comprising the

CC nucleic acid operably linked to a promoter, a recombinant cell
 CC transformed/transfected with the nucleic acid (or expression vector), an
 CC immunogenic composition comprising the nucleic acid in combination with a
 CC pharmaceutical adjuvant or carrier, an isolated S. mansoni protein
 CC comprising an amino acid sequence encoded by the nucleic acid, an
 CC immunogenic composition comprising the isolated S. mansoni protein, an
 CC isolated antibody which specifically binds to the isolated S. mansoni
 CC protein, a hybridoma cell line which produces the isolated antibody, a
 CC method for determining a S. mansoni infection in a subject and a computer
 CC readable medium having recorded in it a nucleic acid molecule from S.
 CC mansoni genome. The nucleic acid molecule is useful for preventing,
 CC diagnosing, or treating S. mansoni infection (schistosomiasis). It can
 CC also be used as vaccine against S. mansoni. The present sequence is a
 CC Schistosoma mansoni protein of the invention. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences. Also, SEQ ID 2141-2152 are
 CC mentioned in the specification but are not included in the sequence
 CC listing.

XX Sequence 352 AA;

Query Match 82.3%; Score 51; DB 9; Length 352;
 Best Local Similarity 90.0%; Pred. No. 5;
 Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
 |||||:
 Db 84 GEKGAGGAPG 93

RESULT 23

ID ABM82963
 ID ABM82963 standard; protein; 675 AA.

XX AC ABM82963;

XX DT 18-NOV-2004 (first entry)

XX DE Human diagnostic and therapeutic pprotein SEQ ID NO:3212.

XX KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX OS Homo sapiens.

XX PN WO2004023973-A2.

XX PD 25-MAR-2004.

XX PF 12-SEP-2003; 2003WO-US028227.

XX PR 12-SEP-2002; 2002US-0410259P.

XX PR 12-SEP-2002; 2002US-0410260P.

XX PA (INCY-) INCYTE CORP.

XX PI Schmidt JP, Wright RJ, Bruns CM, Marianovic MM, Shen F;
 PI Harthshorne TA, Suchorolski MT, Alcius CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;

XX WIPI; 2004-329368/30.
 DR N-PSDB; ACN41615.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.

PS Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX

SQ Sequence 675 AA;

Query Match 80.6%; Score 50; DB 8; Length 675;
Best Local Similarity 81.8%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPL 11
|||||
Db 281 GEGDGSFGI 291

RESULT 24

ABM82962
ID ABM82962 standard; protein; 699 AA.

XX AC ABM82962;

XX 18-NOV-2004 (first entry)

XX Human diagnostic and therapeutic pprotein SEQ ID NO:3211.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
XX N-PSDB; ACN41614.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
XX in diagnosing a condition, disease or disorder associated with human
XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
XX in gene mapping.

XX Claim 27; Page; 190pp; English.

PS

PS Claim 27; Page; 190pp; English.

XX The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorders, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX

SQ Sequence 699 AA;

Query Match 80.6%; Score 50; DB 8; Length 699;
Best Local Similarity 81.8%; Pred. No. 15;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPL 11
|||||
Db 277 GEGDGSFGI 287

RESULT 25

ABM82961
ID ABM82961 standard; protein; 703 AA.

XX AC ABM82961;

XX 18-NOV-2004 (first entry)

XX Human diagnostic and therapeutic pprotein SEQ ID NO:3210.

XX gene therapy; human diagnostic and therapeutic polynucleotide; dithp.

XX Homo sapiens.

XX WO2004023973-A2.

XX 25-MAR-2004.

XX 12-SEP-2003; 2003WO-US028227.

XX 12-SEP-2002; 2002US-0410259P.

XX 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.

XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patry S, Shi X, Suarez CJ;
XX WPI; 2004-329368/30.
XX N-PSDB; ACN41613.

XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
XX in diagnosing a condition, disease or disorder associated with human
XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
XX in gene mapping.

XX Claim 27; Page; 190pp; English.

PS

CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders,
 CC autoimmune/inflammatory disorders, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorder, or
 CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 CC
 XX

SQ Sequence 703 AA;

Query Match 80.6%; Score 50; DB 8; Length 703;
 Best Local Similarity 81.8%; Pred. No. 15;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPL 11
 |||||
 Db 281 GEGKAGSGPI 291

RESULT 26

AA08305
 ID AAY08305 standard; protein; 705 AA.

XX
 AC AAY08305;

XX
 DT 14-JUL-1999 (first entry)

XX
 DE Human collagen IX alpha-2 chain protein.

XX Type IX collagen; alpha-1 chain; alpha-2 chain; alpha-3 chain; human;
 KW detection; autoimmune tissue; connective tissue; disorder; antibody;
 KW antigen-antibody complex; rheumatoid arthritis; osteoarthritis;
 KW diagnosis; chondrosarcoma; endosarcoma; ocular disorder; endochondroma;
 KW chondrodysplasia; cartilage.

XX Homo sapiens.

XX WO921011-A1.

XX 29-APR-1999.

XX 23-OCT-1998; 98WO-US022616.

XX 23-OCT-1997; 97US-0063006P.

XX (FIBR-) FIBROGEN INC.

XX Chunlin Y;

XX WPI; 1999-288409/24.

XX Detecting autoimmune disorders and/or connective tissue disorders
 PT comprises determining levels of type IX collagen in serum.

XX Disclosure; Fig 1B; 36pp; English.

XX This invention describes the detection of an autoimmune/connective tissue
 CC disorder in a patient which comprises (a) reacting a sample with an
 CC antibody specific for type IX collagen (b) allowing (a) to form an
 CC antigen-antibody complex (c) measuring the complex and (d) comparing the
 CC results to a control, non-diseased serum sample to determine presence or
 CC progression of autoimmune/connective tissue disorder. The method is
 CC useful for detecting, diagnosing and monitoring diseases comprising the
 CC abnormal expression of type IX collagen in relevant body fluids or

CC tissues especially rheumatoid arthritis or osteoarthritis. The method is
 CC also useful for detecting or monitoring connective tissue diseases
 CC especially chondrosarcoma, endosarcoma, ocular disorders, endochondroma,
 CC chondrodysplasia and secondary chondrosarcoma. The use of antibodies
 CC provides for early and accurate diagnosis of cartilage related autoimmune
 CC diseases e.g. rheumatoid arthritis or other connective tissue disorders
 XX

SQ Sequence 705 AA;

Query Match 80.6%; Score 50; DB 2; Length 705;
 Best Local Similarity 81.8%; Pred. No. 15;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPL 11
 |||||
 Db 277 GEGKAGSGPI 287

RESULT 27

ADY66076
 ID ADY66076 standard; protein; 232 AA.

XX
 AC ADY66076;

XX 02-JUN-2005 (first entry)

XX S. mansoni protein SEQ ID 1494.

XX Schistosomacide; schistosomiasis; psychiatric disorder; vaccine;
 KW diagnosis.

XX Schistosoma mansoni.

XX WO2005023979-A2.

XX 17-MAR-2005.

XX 10-SEP-2004; 2004WO-BR000170.

XX 11-SEP-2003; 2003US-0502277P.

XX (AMPA-) FUNDACAO AMPARO A PESQUISA DO ESTADO.

XX Verjovski-Almeida S, Leite LCC, Farias LP, Miyasato PA, Kawano T;
 PI Demarco R, Garcia JCL, Martins EAL, Ho PL, Nascimento ALTO;
 PI Dias-Neto E, Setubal JC, Menck CFM, Madeira AMEN, Rodrigues V;
 PI Gargioni C;

XX WPI; 2005-223357/23.

XX New isolated nucleic acid molecule encoding a Schistosoma mansoni
 PT protein, useful for as a vaccine or for preventing, diagnosing, or
 PT treating Schistosoma mansoni infection.

XX Disclosure; SEQ ID NO 1494; 52pp; English.

XX The invention relates to an isolated nucleic acid molecule encoding a
 CC Schistosoma mansoni protein, or its portion which is at least 20 amino
 CC acids in length. Also included are an expression vector comprising the
 CC nucleic acid operably linked to a promoter, a recombinant cell
 CC transformed/transfected with the nucleic acid (for expression vector), an
 CC immunogenic composition comprising the nucleic acid in combination with a
 CC pharmaceutical adjuvant or carrier, an isolated S. mansoni protein
 CC comprising an amino acid sequence encoded by the nucleic acid, an
 CC immunogenic composition comprising the isolated S. mansoni protein, an
 CC isolated antibody which specifically binds to the isolated S. mansoni
 CC protein, a hybridoma cell line which produces the isolated antibody, a
 CC method for determining a S. mansoni infection in a subject and a computer
 CC readable medium having recorded in it a nucleic acid molecule from S.
 CC mansoni genome. The nucleic acid molecule is useful for preventing,
 CC diagnosing, or treating S. mansoni infection (schistosomiasis). It can
 CC also be used as vaccine against S. mansoni. The present sequence is a
 CC Schistosoma mansoni protein of the invention. Note: The sequence data for

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. Also, SEQ ID 2141-2152 are
CC mentioned in the specification but are not included in the sequence
CC listing.
XX
XX
SQ Sequence 232 AA;

Query Match 79.0%; Score 49; DB 9; Length 232;
Best Local Similarity 75.0%; Pred. No. 6.9;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGLL 12
||||:|||||
Db 192 GEKGEGQPGLL 203

RESULT 28
ADF72542
ID ADF72542 standard; protein; 244 AA.
XX
AC ADF72542;
XX
XX 26-FEB-2004 (first entry)
XX Human APM1 OBG3 amino acid sequence.
XX
KW circulating free fatty acid reduction; gOBG3; gOBG3; globular domain;
KW antilipaeamic; antiarteriosclerotic; antidiabetic; hypotensive;
KW gene therapy; body mass reduction; weight loss; obesity-related disorder;
KW hyperlipidaemia; atherosclerosis; diabetes; hypertension; human.
XX
XX Homo sapiens.
OS
XX WO2003102027-A1.
FN
XX 11-DEC-2003.
PD
XX 26-MAY-2003; 2003WO-IB002223.
FF
XX 31-MAY-2002; 2002US-0385238P.
PR
XX (G8ST) GENSET SA.
PA
XX Lucas J, Yeh J, Dialynas D;
PI
XX WPI; 2004-043084/04.
DR
XX
XX Lowering circulating free fatty acid levels for treating obesity-related
XX disorders e.g., hyperlipidemia, atherosclerosis or hypertension by
XX administering a composition comprising a homotrimeric gOBG3 polypeptide
XX fragment.
XX
XX Example 1; Fig 1; 169pp; English.
PS
XX The present invention describes a method for lowering circulating free
XX fatty acid levels in an individual, which comprises administering a
XX composition comprising a carrier and a homotrimeric gOBG3 (OBG3 globular
XX domain) polypeptide fragment. Also described: (1) an isolated
XX homotrimeric gOBG3 polypeptide fragment; (2) a composition comprising a
XX carrier and the gOBG3 polypeptide fragment or polynucleotide or vector;
XX (3) an isolated polynucleotide or its complement encoding the gOBG3
XX polypeptide fragment; (4) a vector comprising the polynucleotide sequence
XX encoding the gOBG3 polypeptide fragment; and (5) a transformed host cell
XX comprising the vector. OBG3 has antilipaeamic, antiarteriosclerotic,
XX antidiabetic and hypotensive activities, and can be used in gene therapy.
XX The method is useful for lowering circulating free fatty acid levels for
XX reducing body mass, maintaining weight loss or for treating obesity-
XX related disorders e.g., hyperlipidaemia, atherosclerosis, diabetes or
XX hypertension. The present sequence is used in the exemplification of the
XX present invention.
XX
XX Sequence 244 AA;

Query Match 79.0%; Score 49; DB 8; Length 244;
Best Local Similarity 75.0%; Pred. No. 7.3;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGLL 12
||||:|||||
Db 63 GEKGEGDGPLL 74

RESULT 29
AAM41798
ID AAM41798 standard; protein; 311 AA.
XX
AC AAM41798;
XX
XX 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6729.
DE
XX
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
XX Homo sapiens.
OS
XX WO200153312-A1.
FN
XX 26-JUL-2001.
PD
XX 26-DEC-2000; 2000WO-US034263.
FF
XX 23-DEC-1999; 99US-00471275.
PR
XX 21-JAN-2000; 2000US-00488725.
PR
XX 25-APR-2000; 2000US-00552317.
PR
XX 20-JUN-2000; 2000US-00598042.
PR
XX 19-JUL-2000; 2000US-00620312.
PR
XX 03-AUG-2000; 2000US-00653450.
PR
XX 14-SEP-2000; 2000US-00662191.
PR
XX 19-OCT-2000; 2000US-00693036.
PR
XX 29-NOV-2000; 2000US-00727344.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR
XX N-PSDB; AAI60954.
DR
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
PT
XX
XX Example 2; SEQ ID NO 6729; 10078pp; English.
PS
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nontropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form

CC part of the printed specification
 XX Sequence 311 AA;
 SQ

Query Match 79.0%; Score 49; DB 4; Length 311;
 Best Local Similarity 72.7%; Pred. No. 9.4;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEGAGSGPGL 11
 ||||:|||||
 Db 289 GEGSGEGPGI 299

RESULT 30
 AAW40113
 ID AAW40113 standard; protein; 549 AA.
 XX AC
 XX AAW40113;
 DT 03-JUN-1998 (first entry)
 XX Human alpha-2(IV) collagen protein.
 DE COL4A6 gene; type IV collagen protein; alpha-6(IV); alpha-2(IV);
 KW Alports syndrome; leiomyomatosis; diagnosis; gene therapy; antibody.
 XX Homo sapiens.
 OS

Key Location/Qualifiers
 FH 167..183
 FT Region
 FT /label= I
 FT /note= "This region interrupts a collagenous repeat"
 FT 235..236
 FT /label= II
 FT /note= "This region interrupts a collagenous repeat"
 FT 255..275
 FT /label= III
 FT /note= "This region interrupts a collagenous repeat"
 FT 291..292
 FT /label= IV
 FT /note= "This region interrupts a collagenous repeat"
 FT 350..360
 FT /label= V
 FT /note= "This region interrupts a collagenous repeat"
 FT 393..400
 FT /label= VI
 FT /note= "This region interrupts a collagenous repeat"
 FT 419..422
 FT /label= VII
 FT /note= "This region interrupts a collagenous repeat"
 FT 484..493
 FT /label= IX
 FT /note= "This region interrupts a collagenous repeat"

US5731192-A.
 XX
 XX 24-MAR-1998.
 PD
 XX 23-JUN-1995; 95US-00494168.
 PF
 XX 27-AUG-1993; 93US-00112465.
 PR
 XX (UYVA) UNIV VALE.
 PA
 XX Zhou J, Reenders ST;
 PI WPI; 1998-216495/19.
 DR Nucleic acid encoding human alpha-6(IV) collagen - useful for, e.g.
 XX diagnosis or gene therapy of Alport's disease.
 PT
 XX Example 1; Fig 3; 43pp; English.
 PS
 XX

CC This sequence is the human alpha-2(IV) collagen protein which is used to
 CC analyse a novel human alpha-6(IV) collagen protein encoded by the COL4A6
 CC gene. The alpha-6(IV) protein can be used for diagnosis or gene therapy
 CC of diseases associated with collagen type IV pathology, especially
 CC Alport's syndrome and associated diffuse leiomyomatosis. The polypeptide
 CC may also be used for generating monoclonal or polyclonal antibodies
 CC having specificity for the alpha-6(IV) polypeptide especially an antibody
 CC that is not crossreactive with other collagen proteins including alpha-
 CC 1(IV), alpha-2(IV) and alpha-5(IV) collagens
 XX
 SQ Sequence 549 AA;

Query Match 79.0%; Score 49; DB 2; Length 549;
 Best Local Similarity 72.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEGAGSGPGL 11
 ||||:|||||
 Db 276 GEGSGEGPGI 286

RESULT 31
 ABM82875
 ID ABM82875 standard; protein; 994 AA.
 XX AC
 XX ABM82875;
 DT 18-NOV-2004 (first entry)
 XX Human diagnostic and therapeutic pprotein SEQ ID NO:3124.
 DE
 XX Human diagnostic and therapeutic pprotein SEQ ID NO:3124.
 KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
 XX Homo sapiens.
 OS
 XX WO2004023973-A2.
 FN
 XX 25-MAR-2004.
 PD
 XX 12-SEP-2003; 2003WO-US028227.
 PP
 XX 12-SEP-2002; 2002US-0410259P.
 PR 12-SEP-2002; 2002US-0410260P.
 XX (INCY-) INCYTE CORP.
 PA
 XX Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthshorne TA, Suchorski MT, Altus CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtson ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patry S, Shi X, Suarez CJ;
 XX
 XX WPI; 2004-329368/30.
 DR N-PSDB; ACN41527.
 DR
 XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 XX in diagnosing a condition, disease or disorder associated with human
 XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 XX in gene mapping.
 PT
 XX Claim 27; Page; 190pp; English.
 PS
 XX The invention relates to novel diagnostic and therapeutic polynucleotides
 XX selected from one of the 2722 sequences defined in the specification. A
 XX polynucleotide of the invention may have a use in gene therapy. The human
 XX diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 XX used to diagnose a particular condition, disease or disorder associated
 XX with human molecules, e.g. cell proliferative disorders,
 XX autoimmune/inflammatory disorder, developmental disorder, endocrine
 XX disorder, neurological disorders, gastrointestinal disorders, or

CC infections caused by virus, bacteria, fungi or parasite. The dithp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 994 AA;

Query Match 79.0%; Score 49; DB 8; Length 994;
 Best Local Similarity 72.7%; Pred. No. 32;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GEGGAGSGPL 11
 ||||:||||:
 Db 276 GEGGSEGEPL 286

RESULT 32
 ABM82874
 ID ABM82874 standard; protein; 1061 AA.
 XX
 AC ABM82874;
 XX
 DT 18-NOV-2004 (first entry)
 XX
 DE Human diagnostic and therapeutic pprotein SEQ ID NO:3123.
 XX
 KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
 XX
 OS Homo sapiens.
 XX
 PN WO2004023973-A2.
 XX
 PD 25-MAR-2004.
 XX

PF 12-SEP-2003; 2003WO-US028227.
 XX
 PR 12-SEP-2002; 2002US-0410259P.
 PR 12-SEP-2002; 2002US-0410260P.
 XX
 PA (INCY-) INCYTE CORP.
 XX
 PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
 PI Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
 PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;
 PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LU;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirtton ES;
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
 PI Patury S, Shi X, Suarez CJ;
 XX
 WPI: 2004-329368/30.
 DR N-PSDB; ACN41526.

XX
 XX New diagnostic and therapeutic polynucleotides and polypeptides, useful
 XX in diagnosing a condition, disease or disorder associated with human
 XX molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 XX in gene mapping.
 XX
 XX Claim 27; Page; 190pp; English.
 PS
 XX The invention relates to novel diagnostic and therapeutic polynucleotides
 XX selected from one of the 2722 sequences defined in the specification. A
 XX polynucleotide of the invention may have a use in gene therapy. The human
 XX diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
 XX used to diagnose a particular condition, disease or disorder associated
 XX with human molecules, e.g. cell proliferative disorders,
 XX autoimmune/inflammatory disorder, developmental disorder, endocrine
 XX disorder, neurological disorders, gastrointestinal disorders, or
 XX infections caused by virus, bacteria, fungi or parasite. The dithp

CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dithp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 1061 AA;

Query Match 79.0%; Score 49; DB 8; Length 1061;
 Best Local Similarity 72.7%; Pred. No. 34;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GEGGAGSGPL 11
 ||||:||||:
 Db 276 GEGGSEGEPL 286

RESULT 33
 AAM40012
 ID AAM40012 standard; protein; 1078 AA.
 XX
 AC AAM40012;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 3157.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US034263.
 XX
 PR 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 PA (HYSE-) HYSEQ INC.

XX
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 XX Zhou P, Goodrich R, Drmanac RT;
 XX
 WPI: 2001-442253/47.
 DR N-PSDB; AAI59168.
 XX

XX Novel nucleic acids and polypeptides, useful for treating disorders such
 XX as central nervous system injuries.

XX Example 4; SEQ ID NO 3157; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 XX encoded polypeptides (AAM38642-AAAM42213) with nootropic,
 XX immunosuppressant and cytostatic activity. The polynucleotides are useful
 XX in gene therapy. A composition containing a polypeptide or polynucleotide
 XX of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification

XX SQ Sequence 1078 AA;

Query Match 79.0%; Score 49; DB 4; Length 1078;
 Best Local Similarity 72.7%; Pred. No. 34;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
 ||||: ||:
 Db 276 GKGSGEGPGI 286

RESULT 34
 AA017361
 ID AA017361 standard; protein; 1712 AA.

XX AC AA017361;

XX DT 19-JUL-2002 (first entry)

XX DE Human alpha-1 type IV collagen.

XX KW Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
 KW aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
 KW insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
 KW transmembrane receptor PTK7; collagen type XVIII alpha 1;
 KW platelet derived growth factor receptor alpha; laminin M chain;
 KW subtilisin like protein PACE4; nidogen.

XX OS Homo sapiens.

XX PN EP1191107-A2.

XX PD 27-MAR-2002.

XX PF 21-AUG-2001; 2001EP-00250300.

XX PR 25-SEP-2000; 2000DE-01048633.

XX PA (SCHD) SCHERING AG.

XX PI Hess-Stumpp H, Haendler B, Kraetzschmar J, Kreft B, Winterhager E;

XX PI Regidor P, Scotti S;

XX DR WPI, 2002-317413/36.

XX PT In vitro diagnosis and monitoring of endometriosis, comprises detecting
 PT reduced expression of specific gene products, e.g. from the fibronectin
 PT gene.

XX Claim 1; Page 16-17; 21pp; German.

XX The present invention relates to a method for the in vitro diagnosis of
 CC endometriosis by determining the amount of gene product from at least one
 CC specific gene in a patient sample and comparing this with the amount of
 CC gene product in a control sample. A reduced level is indicative of
 CC endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
 CC aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
 CC insulin-like growth factor binding protein-2, alpha-2 type IV collagen,
 CC transmembrane receptor PTK7, collagen type XVIII alpha 1, platelet
 CC derived growth factor receptor alpha, laminin M chain, subtilisin like
 CC protein PACE4 or nidogen. The method is useful for initial diagnosis of
 CC endometriosis, and also for monitoring progress and treatment of the

CC disease. The present sequence is human alpha-2 type IV collagen
 XX SQ Sequence 1712 AA;

Query Match 79.0%; Score 49; DB 5; Length 1712;
 Best Local Similarity 72.7%; Pred. No. 56;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
 ||||: ||:
 Db 276 GKGSGEGPGI 286

RESULT 35

ADS10473
 ID ADS10473 standard; protein; 1712 AA.

XX AC ADS10473;

XX DT 16-DEC-2004 (first entry)

XX DE Human therapeutic protein - SEQ ID 710.

XX KW antiinflammatory; neuroprotective; antianaemic; cytostatic; vulnerary;
 KW inflammatory; haematopoiesis; immunity; neurodegenerative; stem cell;
 KW aplastic anaemia; cancer; wound healing; gene therapy.

XX OS Homo sapiens.

XX PN WO2004080148-A2.

XX PD 23-SEP-2004.

XX PF 30-SEP-2003; 2003WO-US030720.

XX PR 02-OCT-2002; 2002US-0416186P.

XX PA (NUVE-) NUVELO INC.

XX PI Tang YT, Asundi V, Ren F, Zhang J, Wehrman T, Wang Z, Ma Y;

XX PI Wang D, Chen R, Zhao QA, Wang J, Ghosh M, Xue AJ, Weng G, Zhou P;

XX DR WPI; 2004-668857/65.

XX DR N-PSDB; ADS09789.

XX PT New polynucleotide, useful in preparing a composition for diagnosing or
 PT treating inflammatory, neurodegenerative or stem cell disorders, e.g.,
 PT aplastic anemia or cancer for promoting wound healing.

XX Claim 20; SEQ ID NO 710; 718pp; English.

XX The invention relates to a novel isolated polynucleotide and the encoded
 CC polypeptide. The molecules of the invention demonstrate antiinflammatory,
 CC neuroprotective, antianaemic, cytostatic and vulnerary activities and may
 CC be useful in preparing a composition for diagnosing or treating
 CC inflammatory, haematopoietic, immune, neurodegenerative or stem cell
 CC disorders, such as aplastic anaemia or cancer, as well as for promoting
 CC wound healing. The molecules may also be utilised during gene therapy
 CC procedures. The current sequence is that of a human therapeutic protein
 CC of the invention. The current sequence is not shown explicitly within the
 CC specification but can be accessed from the WIPO web-site.

XX SQ Sequence 1712 AA;

Query Match 79.0%; Score 49; DB 8; Length 1712;
 Best Local Similarity 72.7%; Pred. No. 56;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
 ||||: ||:
 Db 276 GKGSGEGPGI 286

```
RESULT 36
ADU06526
ID ADU06526 standard; protein; 1712 AA.
XX
AC ADU06526;
XX
DT 27-JAN-2005 (first entry)
XX
DE Novel bronchial cancer-associated human protein SeqID750.
XX
KW bronchial cancer; cytostatic; tumour-associated protein;
KW cancer detection; metastasis; tumour; human.
XX
OS Homo sapiens.
XX
PN DE10316701-A1.
XX
PD 04-NOV-2004.
XX
PF 09-APR-2003; 2003DE-01016701.
XX
PR 09-APR-2003; 2003DE-01016701.
XX
XX (HINZ/) HINZMANN B.
PA (HERM/) HERMANN K.
PA (CAST/) HEIDEN CASTANOS-VELEZ E.
XX
PI Mennerich D, Bruemendorf T, Heiden E, Hermann K, Kinnemann H;
PI Li X, Roepcke S, Staub E, Hinzmann B, Rosenthal A, Pilarsky C;
XX
WPI; 2004-786403/78.
DR N-PSDB; ADU06039.
XX
PT New nucleic acid, and derived proteins, useful for diagnosis of bronchial
PT cancer and in screening for therapeutic and diagnostic agents.
XX
PS Claim 2; SEQ ID NO 750; 1381pp; German.
XX
XX This invention relates to a novel isolated nucleic acid associated with
CC bronchial cancer comprising 489 defined sequences given in the
CC specification. The invention may be useful for the production of
CC compounds with a cytostatic activity through the inhibition of expression
CC or activity of tumour-associated proteins. The novel DNA sequences and
CC the proteins/peptides encoded by them are used for detecting bronchial
CC cancer or determining the risk of developing it and to screen for
CC specific binding partners of the DNA or protein sequences, where the
CC binding partners are potentially useful as agents for treating or
CC diagnosing bronchial cancer. The DNA or protein sequences can also be
CC used for prognosis, detection of metastases and for secondary treatment
CC (of tumours that have been stabilised or are no longer detectable).
CC Detecting abnormal expression of the DNA sequences provides early
CC diagnosis of bronchial cancers. The present sequence is that of a protein
CC encoded by a novel bronchial cancer-associated human gene sequence of the
CC invention.
XX
SQ Sequence 1712 AA;
Query Match 79.0%; Score 49; DB 8; Length 1712;
Best Local Similarity 72.7%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPGL 11
Db 276 GEKGSGEGPGI 286
||||:|||||

RESULT 37
ADY59885
ID ADY59885 standard; protein; 1712 AA.
XX
AC ADY59885;
XX
DT 19-MAY-2005 (first entry)
XX
XX Human collagen IV alpha2 chain.
XX protein interaction; cytostatic; tumor; collagen; cancer; antiangiogenic;
XX breast tumor.
XX
OS Homo sapiens.
XX
PN US2005048063-A1.
XX
PD 03-MAR-2005.
XX
PF 25-AUG-2003; 2003US-00648813.
XX
PR 25-AUG-2003; 2003US-00648813.
XX
PA (RUOS/) RUOSLAHTI E.
PA (ESSL/) ESSLER M.
PA (BROW/) BROWN D M.
XX
PI Ruoslahti E, Essler M, Brown DM;
XX
WPI; 2005-195235/20.
XX
XX Isolated homing peptide or peptidomimetic, useful for directing moiety to
XX tumor vasculature and imaging tumor vasculature in subject.
XX
XX Example 3; SEQ ID NO 4; 41pp; English.
XX
XX The invention relates to an isolated peptide or peptidomimetic (I),
CC comprising a fully defined 5 amino acids sequence (Cys-Arg-Glu-Lys-Ala)
CC given in specification or its peptidomimetic, the peptide or
CC peptidomimetic having a length of less than 100 residues. The peptide (I)
CC can form part of a conjugate (II), comprising a therapeutic agent linked
CC to the homing molecule (I) that selectively homes to tumor vasculature,
CC especially selectively binding collagen. The therapeutic agent is a
CC cancer chemotherapeutic agent, cytotoxic agent, anti-angiogenic agent,
CC polypeptide, nucleic acid molecule or a small molecule. (II) is useful
CC for reducing the number of tumor vessels e.g., breast tumor vessels in a
CC subject, especially for treating cancer e.g., breast cancer. Amino acid
CC sequence binding to the CREKA peptide sequence were isolated by screening
CC a phage display expression library against immobilized CREKA peptide.
CC This sequence corresponds to the human collagen IV alpha2 chain.
XX
SQ Sequence 1712 AA;
Query Match 79.0%; Score 49; DB 9; Length 1712;
Best Local Similarity 72.7%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPGL 11
Db 276 GEKGSGEGPGI 286
||||:|||||

RESULT 38
ABU70813
ID ABU70813 standard; protein; 293 AA.
XX
AC ABU70813;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human adipocyte Selected Interacting domain, SID, #444.
XX
KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;
KW antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.
XX
OS Homo sapiens.
XX
PN WO200286122-A2.
XX
```

```
PD 31-OCT-2002.
XX
XX
XX 14-MAR-2002; 2002WO-EP003768.
XX
XX 14-MAR-2001; 2001US-0275734P.
XX
XX (HYBR-) HYBRIGENICS.
XX
XX Legrain P, Daviet L;
XX
XX WPI; 2003-103412/09.
XX
XX N-PSDB; ACA57357.
XX
XX New complex between two interacting proteins in adipocyte cells, useful
XX for identifying selected interacting domains that modulate protein
XX interactions, or for preventing or treating metabolic disorders such as
XX obesity or diabetes.
XX
XX Claim 6; Page 258-259; 382pp; English.
XX
XX The invention relates to a complex between two interacting proteins in
XX adipocyte cells, given in the specification. The proteins are identified
XX by selecting a bait protein from a known adipocyte marker and then
XX performing a yeast 2-hybrid selection to isolate prey proteins encoded by
XX members of an adipocyte cDNA library. The proteins are designated SID
XX (RTM) (selected interacting domains) proteins. Also included are a
XX polynucleotide encoding a polypeptide in the adipocyte cells, a
XX recombinant host cell expressing at least one of the interacting
XX polypeptides of the complex, selecting a modulating compound in adipocyte
XX cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
XX sequences given in the specification (including its fragment or variant),
XX a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
XX given in the specification (including its fragment or variant), a vector
XX comprising the SID (RTM) polynucleotide, a recombinant host cell
XX comprising the vector, a protein chip comprising the polypeptides and a
XX record comprising all or part of the data, listed in the specification.
XX The complex, polypeptides, polynucleotides and compounds are useful for
XX preventing or treating metabolic disorders such as obesity or diabetes.
XX The polynucleotides are useful as probes or primers. The complex is
XX particularly useful for identifying selected interacting domains (SID
XX (RTM)) for screening drugs that modulate the protein interaction, thus
XX exhibiting the therapeutic effect. The present sequence represents a SID
XX (prey) protein of the invention
XX
XX Sequence 293 AA;
SQ
Query Match 77.4%; Score 48; DB 6; Length 293;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 GEGKAGSGSPG 10
Db 101 GERGSGSGPG 110
||:|||||
||:|||||
RESULT 39
AAU23675
ID AAU23675 standard; protein; 309 AA.
XX
XX AAU23675;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #761.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
XX ligase; hyperproliferative disorder; immunodeficiency disorder;
XX autoimmune disorder; neurological disorder; metabolic disorder;
XX inflammatory disorder; cardiovascular disorder; reproductive disorder;
XX blood-related disorder; infectious disorder; cytostatic; anti arthritic;
XX nephrotropic; anticoagulant.
XX
XX Homo sapiens.
OS
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001239.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX
XX 04-FEB-2000; 2000US-0180628P.
XX
XX 24-FEB-2000; 2000US-0184664P.
XX
XX 02-MAR-2000; 2000US-0186350P.
XX
XX 16-MAR-2000; 2000US-0189874P.
XX
XX 17-MAR-2000; 2000US-0190076P.
XX
XX 18-APR-2000; 2000US-0198123P.
XX
XX 19-MAY-2000; 2000US-0205515P.
XX
XX 07-JUN-2000; 2000US-0209467P.
XX
XX 28-JUN-2000; 2000US-0214886P.
XX
XX 30-JUN-2000; 2000US-0215135P.
XX
XX 07-JUL-2000; 2000US-0216647P.
XX
XX 07-JUL-2000; 2000US-0216880P.
XX
XX 11-JUL-2000; 2000US-0217487P.
XX
XX 14-JUL-2000; 2000US-0218290P.
XX
XX 26-JUL-2000; 2000US-0220963P.
XX
XX 14-AUG-2000; 2000US-0224518P.
XX
XX 14-AUG-2000; 2000US-0224519P.
XX
XX 14-AUG-2000; 2000US-0225213P.
XX
XX 14-AUG-2000; 2000US-0225214P.
XX
XX 14-AUG-2000; 2000US-0225266P.
XX
XX 14-AUG-2000; 2000US-0225267P.
XX
XX 14-AUG-2000; 2000US-0225268P.
XX
XX 14-AUG-2000; 2000US-0225270P.
XX
XX 14-AUG-2000; 2000US-0225447P.
XX
XX 14-AUG-2000; 2000US-0225757P.
XX
XX 14-AUG-2000; 2000US-0225758P.
XX
XX 14-AUG-2000; 2000US-0225759P.
XX
XX 18-AUG-2000; 2000US-0226279P.
XX
XX 22-AUG-2000; 2000US-0226681P.
XX
XX 22-AUG-2000; 2000US-0226868P.
XX
XX 22-AUG-2000; 2000US-0227182P.
XX
XX 23-AUG-2000; 2000US-0227009P.
XX
XX 30-AUG-2000; 2000US-0228924P.
XX
XX 01-SEP-2000; 2000US-0229287P.
XX
XX 01-SEP-2000; 2000US-0229343P.
XX
XX 01-SEP-2000; 2000US-0229344P.
XX
XX 01-SEP-2000; 2000US-0229345P.
XX
XX 05-SEP-2000; 2000US-0229509P.
XX
XX 05-SEP-2000; 2000US-0229513P.
XX
XX 06-SEP-2000; 2000US-0230437P.
XX
XX 06-SEP-2000; 2000US-0230438P.
XX
XX 08-SEP-2000; 2000US-0231242P.
XX
XX 08-SEP-2000; 2000US-0231243P.
XX
XX 08-SEP-2000; 2000US-0231244P.
XX
XX 08-SEP-2000; 2000US-0231413P.
XX
XX 08-SEP-2000; 2000US-0231414P.
XX
XX 08-SEP-2000; 2000US-0232080P.
XX
XX 08-SEP-2000; 2000US-0232081P.
XX
XX 12-SEP-2000; 2000US-0231968P.
XX
XX 14-SEP-2000; 2000US-0232397P.
XX
XX 14-SEP-2000; 2000US-0232398P.
XX
XX 14-SEP-2000; 2000US-0232399P.
XX
XX 14-SEP-2000; 2000US-0232400P.
XX
XX 14-SEP-2000; 2000US-0232401P.
XX
XX 14-SEP-2000; 2000US-0233063P.
XX
XX 14-SEP-2000; 2000US-0233064P.
XX
XX 21-SEP-2000; 2000US-0233065P.
XX
XX 21-SEP-2000; 2000US-0234223P.
XX
XX 21-SEP-2000; 2000US-0234274P.
XX
XX 25-SEP-2000; 2000US-0234997P.
XX
XX 25-SEP-2000; 2000US-0234998P.
XX
XX 26-SEP-2000; 2000US-0235484P.
XX
XX 27-SEP-2000; 2000US-0235834P.
XX
```


PR 07-JUN-2000; 2000US-0209467P.
PR 14-SEP-2000; 2000US-0232398P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251990P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-476195/51.
XX N-PSDB; ABK72051.
XX Novel isolated human ovarian related polypeptide useful for
XX diagnosis/treatment of disorders of ovary and breast such as neoplastic
XX disorders, infectious diseases, inflammatory diseases, and reproductive
XX disorders.
XX Claim 11; SEQ ID NO 78; 524pp; English.
XX
XX The invention relates to isolated ovarian related polypeptide (ovarian
XX antigen) comprising a sequence at least 90% identical to a sequence
XX selected from a polypeptide fragment, domain, epitope or full length
XX protein of a sequence (S1) appearing as ABG60239-ABG60296 having
XX biological activity, or a variant, allelic variant or species homologue
XX of S1. Also included are the cDNA clones encoding the proteins of S1. S1,
XX an anti-S1 antibody and the cDNA are useful for diagnosing, preventing,
XX treating or ameliorating a medical condition in mammalian subject
XX especially diseases and/or disorders of the ovary and/or breast such as
XX neoplastic diseases (such as ovarian Krukenberg tumour and cancer),
XX infectious diseases (e.g., mastitis, oophoritis), inflammatory diseases
XX (e.g., abscesses), reproductive system disorders (Paget's disease),
XX autoimmune disorders (systemic lupus erythematosus, rheumatoid
XX arthritis), blood-related disorders (sickle cell anaemia),
XX hyperproliferative disorders, urinary system disorders
XX (glomerulonephritis), cardiovascular disorders (arrhythmias), respiratory
XX disorders, musculoskeletal system disorders, neural activity and
XX neurological disorders (Alzheimer's disease and Parkinson's disease),
XX endocrine disorders (Addison's disease), gastrointestinal disorders
XX (inflammatory disorders), liver disorders (biliary liver cirrhosis),
XX pancreatic and gall bladder disorders, disorders of the large intestine,
XX developmental and inherited disorders, diseases at the cellular level,
XX and wound healing and epithelial cell proliferation. They are also useful
XX to prevent skin aging, for preventing hair loss, to maintain organs
XX before transplantation or for supporting cell culture of primary tissues,
XX to modulate mammalian characteristics such as body height, to modulate
XX mammalian metabolism, to change a mammal's mental or physical state, and
XX as food additive or preservative. The present sequence represents an
XX ovarian antigen, S1 protein of the invention. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 309 AA;
XX
XX Query Match 77.4%; Score 48; DB 4; Length 309;
XX Best Local Similarity 80.0%; Pred. No. 14;
XX Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GKGAGGSPG 10
XX ||:|||||
XX Db 13 GERSEGGSPG 22
XX
XX RESULT 41
XX ABG61719
XX ID ABG61719 standard; protein; 309 AA.
XX
XX AC ABG61719;

XX 26-AUG-2002 (first entry)
XX Novel ovarian related polypeptide #10.
XX Ovarian related polypeptide; neoplastic disorder; tumour; ovarian cancer;
XX hyperproliferative disorder; adult acute lymphocytic leukaemia;
XX breast cancer; reproductive system disorder; tuberculosis; arthritis;
XX immune system disorder; Chediak-Higashi's syndrome; neonatal neutropenia;
XX autoimmune disorder; Hashimoto's thyroiditis; inflammatory disorder;
XX septic shock; multiple sclerosis; central nervous system disorder;
XX neurological disorder; allergy; Parkinson's disease; Alzheimer's disease;
XX cardiovascular disorder; atherosclerosis; blood related disorder;
XX respiratory disorder; urinary system disorder; musculoskeletal disorder;
XX osteoporosis; wound healing; endocrine disorder; infectious disease;
XX gastrointestinal disorder; transplantation; food additive; preservative.
XX Homo sapiens.
XX
XX US2002045230-A1.
XX
XX 18-APR-2002.
XX
XX 20-JUL-2001; 2001US-00908711.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225267P.
XX 14-AUG-2000; 2000US-0225268P.
XX 14-AUG-2000; 2000US-0225270P.
XX 14-AUG-2000; 2000US-0225447P.
XX 14-AUG-2000; 2000US-0225757P.
XX 14-AUG-2000; 2000US-0225758P.
XX 14-AUG-2000; 2000US-0225759P.
XX 18-AUG-2000; 2000US-0226279P.
XX 22-AUG-2000; 2000US-0226681P.
XX 22-AUG-2000; 2000US-0226868P.
XX 22-AUG-2000; 2000US-0227182P.
XX 23-AUG-2000; 2000US-0227009P.
XX 30-AUG-2000; 2000US-0228924P.
XX 01-SEP-2000; 2000US-0229287P.
XX 01-SEP-2000; 2000US-0229343P.
XX 01-SEP-2000; 2000US-0229344P.
XX 01-SEP-2000; 2000US-0229345P.
XX 05-SEP-2000; 2000US-0229509P.
XX 05-SEP-2000; 2000US-0229513P.
XX 06-SEP-2000; 2000US-0230437P.
XX 06-SEP-2000; 2000US-0230438P.
XX 08-SEP-2000; 2000US-0231242P.
XX 08-SEP-2000; 2000US-0231243P.
XX 08-SEP-2000; 2000US-0231244P.
XX 08-SEP-2000; 2000US-0231413P.

fibrotic disease; human; chimeric protein.

Homo sapiens.

WO9708311-A1.

06-MAR-1997.

30-AUG-1996; 96WO-GB002122.

31-AUG-1995; 95GB-00017773.

23-MAR-1996; 96GB-00006152.

14-JUN-1996; 96GB-00012476.

(UTMA-) UNIV VICTORIA MANCHESTER.

Bulleid N, Kadler K;

WPI; 1997-179268/16.

Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached to an alien collagen alpha-chain or non-collagen material, useful e.g. for wound healing.

Example 1; Page 32-35; 69pp; English.

This sequence represents a chimeric procollagen molecule of the invention. This sequence has the procollagen C-propeptide from the pro-alpha2(I) chain attached to the pro-alpha1(III) chain sequence. The C-propeptide is implicated in the assembly of the monomer chains into trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides and formation of collagen in fibril-forming pro-alpha chains. The C-propeptides determine the type-specific assembly of the moieties to which they are attached. The molecule of the invention comprises a first moiety having procollagen C-propeptide activity attached to a second moiety, which is an alien collagen alpha-chain or a non-collagen material. The novel collagen molecule can be used for treatment or diagnosis in humans or animals, especially for the treatment of procollagen suicide, as an adhesive or implant, to promote (chronic) wound healing or fibrotic diseases with reduced scarring or for use in photography, brewing, foodstuffs or textiles. The novel collagen molecules, especially when containing substitutions in the recognition site, may have significantly altered properties and characteristics, such as different binding kinetics or alpha-chain selection properties

Sequence 623 AA;

Query Match 77.4%; Score 48; DB 2; Length 623;

Best Local Similarity 80.0%; Pred. No. 28;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGSPG 10
||:|:|:|:|

DB 330 GERGSGSGSPG 339

RESULT 44

AAW12842

ID AAW12842 standard; protein; 626 AA.

AC AAW12842;

XX

XX 15-DEC-1997 (first entry)

XX Truncated pro-alpha(III) chain.

XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;

XX trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;

XX fibrotic disease; human.

XX Homo sapiens.

XX WO9708311-A1.

yeast 2-hybrid assay; metabolic disorder; obesity.

Homo sapiens.

WO200286122-A2.

31-OCT-2002.

14-MAR-2002; 2002WO-EP003768.

14-MAR-2001; 2001US-0275734P.

(HYBR-) HYBRIGENICS.

Legrain P, Daviet L;

WPI; 2003-103412/09.

N-PSDB; ACA57279.

New complex between two interacting proteins in adipocyte cells, useful for identifying selected interacting domains that modulate protein interactions, or for preventing or treating metabolic disorders such as obesity or diabetes.

Claim 6; Page 229; 382pp; English.

The invention relates to a complex between two interacting proteins in adipocyte cells, given in the specification. The proteins are identified by selecting a bait protein from a known adipocyte marker and then performing a yeast 2-hybrid selection to isolate prey proteins encoded by members of an adipocyte cDNA library. The proteins are designated SID (RTM) (selected interacting domains) proteins. Also included are a polynucleotide encoding a polypeptide in the adipocyte cells, a recombinant host cell expressing at least one of the interacting polypeptides of the complex, selecting a modulating compound in adipocyte cells, a SID (RTM) polypeptide comprising any of the 738 amino acid sequences given in the specification (including its fragment or variant), a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences given in the specification (including its fragment or variant), a vector comprising the SID (RTM) polynucleotide, a recombinant host cell comprising the vector, a protein chip comprising the polypeptides and a record comprising all or part of the data, listed in the specification. The complex, polypeptides, polynucleotides and compounds are useful for preventing or treating metabolic disorders such as obesity or diabetes. The polynucleotides are useful as probes or primers. The complex is particularly useful for identifying selected interacting domains (SID (RTM)) for screening drugs that modulate the protein interaction, thus exhibiting the therapeutic effect. The present sequence represents a SID (prey) protein of the invention

Sequence 339 AA;

Query Match 77.4%; Score 48; DB 6; Length 339;

Best Local Similarity 80.0%; Pred. No. 15;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGSPG 10
||:|:|:|:|

DB 43 GERGSGSGSPG 52

RESULT 43

AAW12843

ID AAW12843 standard; peptide; 623 AA.

XX

XX AAW12843;

XX 15-DEC-1997 (first entry)

XX Pro-alpha(III):(I) CP chimeric protein.

XX C-propeptide; recognition sequence; procollagen; monomer chain; therapy;

XX trimerised pro-alpha chain; fibril; procollagen suicide; wound healing;

```

XX PD 06-MAR-1997.
XX PF 30-AUG-1996; 96WO-GB002122.
XX PR 31-AUG-1995; 95GB-00017773.
XX PR 23-MAR-1996; 96GB-00006152.
XX PR 14-JUN-1996; 96GB-00012476.
XX PA (UYWA-) UNIV VICTORIA MANCHESTER.
XX PI Bulleid N, Kadler K;
XX DR WPI, 1997-179268/16.
XX DR N-PSDB; AAT59892.
XX PT Novel pro:collagen mol. - comprising pro:collagen C-pro:peptide attached
XX PT to an alien collagen alpha-chain or non-collagen material, useful e.g.
XX PT for wound healing.
XX PS Example 1; Page 28-31; 69pp; English.
XX CC This sequence represents a truncated procollagen pro-alpha(III) chain
XX CC that can be used in the procollagen molecules of the invention. The C-
XX CC propeptide is implicated in the assembly of the monomer chains into
XX CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides
XX CC and formation of collagen in fibril-forming pro-alpha chains. The C-
XX CC propeptides determine the type-specific assembly of the moieties to which
XX CC they are attached. The molecule of the invention comprises a first moiety
XX CC having procollagen C-propeptide activity attached to a second moiety,
XX CC which is an alien collagen alpha-chain or a non-collagen material. The
XX CC novel collagen molecule can be used for treatment or diagnosis in humans
XX CC or animals, especially for the treatment of procollagen suicide, as an
XX CC adhesive or implant, to promote (chronic) wound healing or fibrotic
XX CC diseases with reduced scarring or for use in photography, brewing,
XX CC foodstuffs or textiles. The novel collagen molecules, especially when
XX CC containing substitutions in the recognition site, may have significantly
XX CC altered properties and characteristics, such as different binding
XX CC kinetics or alpha-chain selection properties
XX SQ Sequence 626 AA;
    Query Match 77.4%; Score 48; DB 2; Length 626;
    Best Local Similarity 80.0%; Pred. No. 29;
    Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGGAEGSPG 10
Db ||:|||||
330 GERGEGSGSPG 339

RESULT 45
ADN35278
ID ADN35278 standard; protein; 1028 AA.
XX AC ADN35278;
XX DT 01-JUL-2004 (first entry)
XX DE Helical domain of type III collagen #2.
XX KW tissue sealant; synthetic gelatin; Vunlerary; surgical incision;
XX KW angioplasty; laceration; disease transmission.
XX OS Homo sapiens.
XX PN WO2004028404-A2.
XX PD 08-APR-2004.
XX PF 30-SEP-2003; 2003WO-US031006.
XX PR 30-SEP-2002; 2002US-0415309P.

XX PA (FIBR-) FIBROGEN INC.
XX PI Yang C, Chang RC, Olsen DR, Polarek JW;
XX DR WPI; 2004-375406/35.
XX PT Tissue sealant composition comprises cross linking agent and a synthetic
XX PT collagen or a synthetic gelatin in a dry state.
XX PS Disclosure; SEQ ID NO 2; 75pp; English.
XX CC The present invention relates to a tissue sealant composition comprises a
XX CC crosslinking agent and a synthetic collagen or a synthetic gelatin in a
XX CC dry state, they react with each other upon contact with an environment
XX CC comprising a physiological pH to form a compound. The compound is useful
XX CC to seal a wound comprising a surgical incision (comprising an
XX CC angioplasty), laceration or a puncture and contains controlled and
XX CC reproducible materials with minimal or no risk of infectivity and disease
XX CC transmission. The present sequence represents the helical domain of type
XX CC III collagen.
XX SQ Sequence 1028 AA;
    Query Match 77.4%; Score 48; DB 8; Length 1028;
    Best Local Similarity 80.0%; Pred. No. 48;
    Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGGAEGSPG 10
Db ||:|||||
1003 GERGEGSGSPG 1012

RESULT 46
AAR71704
ID AAR71704 standard; protein; 1078 AA.
XX AC AAR71704;
XX DT 25-MAR-2003 (revised)
XX DT 17-OCT-1995 (first entry)
XX DE Collagen alpha 1 (III) chain precursor.
XX KW Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;
XX KW disorder; osteoporosis; metastatic progression; Paget's disease;
XX KW hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis;
XX KW vasculitis syndrome.
XX OS Homo sapiens.
XX PN WO9508115-A1.
XX PD 23-MAR-1995.
XX PF 19-SEP-1994; 94WO-DK000348.
XX PR 17-SEP-1993; 93DK-00001040.
XX PA (OSTE-) OSTEOMETER AS.
XX PI Qvist P, Bonde M;
XX DR WPI; 1995-131456/17.
XX PT Assaying collagen fragments in body fluid by immunoassay - using
XX PT antibodies raised against synthetic peptide(s) contg. potential
XX PT crosslinking sites, to diagnose and monitor disorders of collagen
XX PT metabolism, e.g. osteoporosis.
XX PS Disclosure (Appendix A); Page 55; 87pp; English.
XX CC Determination of collagen fragments in body fluids can be achieved by

```

CC immunoassay using antibodies directed against synthetic peptides derived
 CC from collagen which contain sites of potential crosslinking. The method
 CC is used to diagnose and monitor treatment of disorders of collagen
 CC metabolism (degradation of type I collagen may indicate osteoporosis,
 CC metastatic progression, Paget's disease, hyperthyroidism or other
 CC conditions involving excessive bone resorption; degradation of type II
 CC collagen may indicate rheumatoid arthritis or osteoarthritis; and of type
 CC III collagen, vacuolitis syndrome). The method can also be used to assess
 CC the toxicity of a compound and to test drugs for their effect on collagen
 CC metabolism. (Updated on 25-MAR-2003 to correct FN field.)
 XX
 SQ Sequence 1078 AA;

Query Match 77.4%; Score 48; DB 2; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGSPG 10
 DB 1023 GERGEGSGSPG 1032
 ||:|||||

RESULT 47
 AAY96125
 ID AAY96125 standard; peptide; 1078 AA.
 XX

AC AAY96125;

XX 19-DEC-2000 (first entry)

DT Collagen type III alpha-1.

XX Collagen type III; vasculitis syndrome; assay; diagnosis.

XX Homo sapiens.

XX US6110689-A.

XX 29-AUG-2000.

XX 04-NOV-1997; 97US-00963825.

XX 21-JAN-1994; 94US-00187319.

XX (OSTE-) OSTEOMETER AS.

XX Bonde M, Qvist P;

XX WPI; 2000-586349/55.

XX Assaying type I collagen fragments for diagnosing osteoporosis in
 PT postmenopausal woman, involves contacting body fluid with synthetic
 PT collagen peptide and antibody and quantifying by competitive binding
 PT assay.

XX Disclosure; Col 46-51; 41pp; English.

XX The present sequence is that of human type III collagen alpha-1. The
 CC invention is based on the discovery of the presence of particular
 CC collagen fragments in body fluids of patients compared with those of
 CC healthy subjects. These fragments are generated upon collagen degradation
 CC and are partly characterised by the presence of potential sites for
 CC crosslinking. A method for assaying collagen fragments in a body fluid
 CC sample is based on the competitive binding to immunological binding
 CC partners of collagen fragments in the sample and of synthetic peptides
 CC derived from collagen and containing crosslinkable sites (see AAY96118-
 CC 21). When considering the degradation of type III collagen, the assay can
 CC be used as a means of identifying the presence of vasculitis syndrome
 XX

Query Match 77.4%; Score 48; DB 3; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 50;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEGGAGSGSPG 10
 DB 1023 GERGEGSGSPG 1032
 ||:|||||

RESULT 48
 AAE16478
 ID AAE16478 standard; protein; 1078 AA.
 XX

AC AAE16478;

XX 09-APR-2002 (first entry)

XX Human collagen alpha (III) protein.

XX Human; collagen; osteoporosis; bone resorption; Paget's disease;
 KW hyperparathyroidism; rheumatoid arthritis; osteoarthritis; therapy;
 KW vasculitis syndrome; toxic substance; tissue degradation; alpha (III).
 XX

XX Homo sapiens.

XX US6323314-B1.

XX 27-NOV-2001.

XX 10-FEB-2000; 2000US-00500811.

XX 21-JAN-1994; 94US-00187319.

XX 04-NOV-1997; 97US-00963825.

XX (OSTE-) OSTEOMETER AS.

XX Qvist P, Bonde M;

XX WPI; 2002-096598/13.

XX Novel peptides that match alpha 1 or 2(I) telopeptide component of cross-

PT linked telopeptide degradation product of type I collagen useful for

PT diagnosing disorders associated with collagen metabolism e.g.

PT osteoporosis.

XX Disclosure; Col 43-50; 34pp; English.

XX The invention relates to a peptide synthesised to match an alpha(I) or
 CC alpha2(I) telopeptide component of a cross-linked telopeptide degradation
 CC product of type I collagen. The method is useful for assaying collagen
 CC fragments in animal body fluids, for determining the degradation of human
 CC collagen types I, II and III and for diagnosing the presence of disorders
 CC associated with the metabolism of collagen, especially osteoporosis. The
 CC peptide is used for assessing the impact of drugs on collagen metabolism.
 CC The peptide is useful in methods to assess an abnormal condition of a
 CC subject for e.g. excessive bone resorption which shows the presence of an
 CC osteoporotic condition or the metastatic progress of a malignancy,
 CC Paget's disease and hyperparathyroidism. Disease states involving
 CC connective tissues can be monitored by determining collagen degradation,
 CC examples are collagen type II degradation associated with rheumatoid
 CC arthritis, osteoarthritis, and collagen type III degradation in
 CC vasculitis syndrome. Since the conditions of the subject can be monitored
 CC continuously, application of these assays can also be used to monitor the
 CC progress of therapy administered to treat these or other conditions and
 CC as a measure of toxicity, since the administration of toxic substances
 CC often results in tissue degradation. The present sequence is human
 CC collagen alpha (III) protein used in the invention
 XX

SQ Sequence 1078 AA;

Query Match 77.4%; Score 48; DB 5; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGSPG 10

Db 1023 GERSESGSPG 1032

RESULT 49
ABB0736
ID ABB0736 standard; protein; 1078 AA.
XX
AC ABB0736;
XX
DT 15-JUL-2002 (first entry)
XX
DE Collagen type III-alpha protein.
XX
KW Collagen; osteoarthritis; Paget's disease; Marfan syndrome; dwarfism;
KW osteogenesis imperfecta; neoplastic growth; rheumatoid arthritis;
KW vasculitis; collagen type III-alpha.
XX
OS Homo sapiens.
XX
PN US6355442-B1.
XX
PD 12-MAR-2002.
XX
PF 13-APR-2000; 2000US-00548608.
XX
PR 21-JAN-1994; 94US-00187319.
XX
FR 04-NOV-1997; 97US-00963825.
XX
PA (OSTE-) OSTEOMETER BIOTECH AS.
XX
PI Qvist P, Bonde M;
XX
XX WPI; 2002-380937/41.
XX
XX Assaying type I collagen fragments in body fluid, useful for diagnosis
PT and assessing treatment of e.g. osteoarthritis, by competitive
PT immunoassay.
XX
PS Disclosure; Col 45-50; 35pp; English.
XX
XX The invention relates to a method for assaying type I collagen fragments
CC (I) in body fluid. The method involves treating the test sample with: (i)
CC synthetic peptide, immobilised on a support; and (ii) immunological
CC binding partner, reactive with the synthetic peptide, so that (i) and the
CC synthetic peptide compete for binding, and (i) are quantified by
CC measuring the binding of the binding partner to the synthetic peptide.
CC The method is used to diagnose disorders of collagen metabolism,
CC especially osteoarthritis but also Paget's disease, Marfan syndrome,
CC osteogenesis imperfecta, neoplastic growth of collagenous tissue,
CC dwarfism, rheumatoid arthritis or vasculitis, also for clinical testing
CC of drugs to assess their effect on collagen metabolism. The present
CC sequence represents the collagen type III-alpha protein
XX
SQ Sequence 1078 AA;
Query Match 77.4%; Score 48; DB 5; Length 1078;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEKGAGSGP 10
||:|||||
Db 1023 GERSESGSPG 1032

RESULT 50
ABB09628
ID ABB09628 standard; peptide; 1078 AA.
XX
AC ABB09628;
XX
DT 29-MAY-2002 (first entry)
XX
DE Human collagen alpha(III) chain precursor.
XX

DE Amino acid sequence of human collagen type III alpha1.
XX
KW Collagen; bone resorption; collagen metabolism; Paget's disease;
KW Marfan's syndrome; osteogenesis imperfecta; neoplastic growth; dwarfism;
KW rheumatoid arthritis; osteoarthritis; vasculitis syndrome.
XX
OS Homo sapiens.
XX
PN US6342361-B1.
XX
PD 29-JAN-2002.
XX
PF 12-MAY-2000; 2000US-00570573.
XX
PR 21-JAN-1994; 94US-00187319.
XX
FR 04-NOV-1997; 97US-00963825.
XX
PA (OSTE-) OSTEOMETER BIOTECH AS.
XX
PI Qvist P, Bonde M;
XX
XX WPI; 2002-224940/28.
XX
XX Assaying type II collagen fragments in a body fluid sample, for
PT determining abnormalities in collagen metabolism, e.g. in Paget's disease
PT or Marfan's syndrome, comprises contacting the fluid with a synthetic
PT type II collagen alpha.
XX
PS Disclosure; Col 45-50; 35pp; English.
XX
XX The present sequence represents human collagen type III alpha1. The
CC specification describes a method for assaying type II collagen fragments
CC in a body fluid sample. The method comprises contacting the body fluid
CC with a synthetic peptide consisting essentially of a type II collagen
CC alpha1 amino acid sequence that competes with the collagen fragments for
CC binding with the immunological binding partner. The method is useful for
CC assaying collagen fragments in body fluids. The method is particularly
CC useful in assays for measuring bone resorption rates in patients or in
CC healthy subjects. The method may be used for determining abnormalities in
CC collagen metabolism, especially in Paget's disease, Marfan's syndrome,
CC osteogenesis imperfecta, neoplastic growth in collagenous tissue,
CC dwarfism, rheumatoid arthritis, osteoarthritis or vasculitis syndrome.
CC The method may be used for determining the degradation of human collagen
CC of type I, II and III. It also can be used during clinical testing of new
CC drugs to assess the impact of these drugs on collagen metabolism. The
CC assays, can be used as a measure of toxicity, since the administration of
CC toxic substances often results in tissue degradation. Thus, the assays
CC may be applied in any situation where the metabolic condition of collagen
CC tissues can be used as an index of the condition, treatment or effect of
CC substances directly administered to the subject or to which the subject
CC is exposed in the environment
XX
SQ Sequence 1078 AA;
Query Match 77.4%; Score 48; DB 5; Length 1078;
Best Local Similarity 80.0%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEKGAGSGP 10
||:|||||
Db 1023 GERSESGSPG 1032

RESULT 51
ADF13078
ID ADF13078 standard; protein; 1078 AA.
XX
AC ADF13078;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human collagen alpha(III) chain precursor.
XX

KW collagen fragment; collagen; collagen metabolism disorder;
 KW collagen degradation; crosslinking site; human; alpha1(III) chain.
 XX Homo sapiens.
 OS
 XX
 XX
 XX
 XX US2003119058-A1.
 XX
 XX 26-JUN-2003.
 XX
 XX 29-JAN-2002; 2002US-00058124.
 XX
 XX 21-JAN-1994; 94US-00187319.
 PR 04-NOV-1997; 97US-00963825.
 PR 12-MAY-2000; 2000US-00570573.
 XX
 XX (OSTE-) OSTEOMETER AS.
 XX
 XX Qvist P, Bonde M;
 XX
 XX WPI; 2003-897106/82.
 XX
 XX Determining collagen fragments in a body fluid sample using antibodies
 PT against synthetic peptides containing crosslinking sites is useful to
 PT diagnose collagen metabolism disorders or if a subject suffers from
 PT collagen degradation.
 XX
 XX Example 3; SEQ ID NO 21; 41pp; English.
 XX
 XX The invention relates to a new method for determining collagen fragments
 CC in a body fluid which comprises contacting the sample with an
 CC immunological binding partner for the fragments, where the binding
 CC partner is immunoreactive with synthetic peptides whose sequences are
 CC essentially derived from collagen and contain potential sites for cross
 CC linking, and is incorporated as a whole antibody or its immunological
 CC fragment in an assay for quantitative determination of collagen fragments
 CC in the sample. The invention is useful for diagnosing disorders
 CC associated with the metabolism of collagen or whether a susceptible
 CC subject is suffering from collagen degradation. A test kit for
 CC quantitating the amount of collagen fragments in a body fluid is claimed.
 CC The present sequence represents the amino acid sequence of human collagen
 CC alpha1(III) chain precursor.
 XX
 XX Sequence 1078 AA;
 SQ
 Query Match 77.4%; Score 48; DB 7; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 50;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEGGAGGSPG 10
 DB 1023 GERGSEGSFG 1032
 RESULT 52
 AAR28916
 ID AAR28916 standard; protein; 1196 AA.
 XX
 AC AAR28916;
 XX
 XX 25-MAR-2003 (revised)
 DT 24-MAR-1993 (first entry)
 XX
 XX Type III procollagen (prior art).
 DE
 XX Mutation; pro-alpha1(III); primer; PCR.
 KW
 XX Homo sapiens.
 OS
 XX
 XX WO9219754-A1.
 PN
 XX 12-NOV-1992.
 PD
 XX 08-MAY-1992; 92WO-US003866.
 PF

XX 08-MAY-1991; 91US-00696607.
 XX
 XX (UVJE-) UNIV JEFFERSON THOMAS.
 PA
 XX Prockop DJ, Tromp GC, Kuivaniemi SH;
 FI
 XX WPI; 1992-398878/48.
 DR
 XX N-PSDB; AAQ30849.
 DR
 XX Kit for detecting genetic pre-disposition for vascular aneurysms -
 PT contains primer 70 amplify portions of Type III procollagen DNA and
 PT detects mutation in standard procollagen DNA.
 XX
 XX Disclosure; Fig 1A-F; 44pp; English.
 XX
 XX Example 1 describes the determination of the presence of a mutation in
 CC the pro-alpha1(III) gene. Primers used in PCR are given in AAQ30834-48.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX Sequence 1196 AA;
 SQ
 Query Match 77.4%; Score 48; DB 2; Length 1196;
 Best Local Similarity 80.0%; Pred. No. 56;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GEGGAGGSPG 10
 DB 1170 GERGSEGSFG 1179
 RESULT 53
 ADN35279
 ID ADN35279 standard; protein; 1313 AA.
 XX
 AC ADN35279;
 XX
 XX 01-JUL-2004 (first entry)
 DT
 XX Synthetic collagen used as sealant tissue protein.
 DE
 XX tissue sealant; synthetic gelatin; Vulerary; surgical incision;
 KW angioplasty; laceration; disease transmission.
 KW
 XX Synthetic.
 OS
 XX WO2004028404-A2.
 PN
 XX 08-APR-2004.
 PD
 XX 30-SEP-2003; 2003WO-US031006.
 PF
 XX 30-SEP-2002; 2002US-0415309P.
 PR
 XX (FIBR-) FTBROGEN INC.
 PA
 XX Yang C, Chang RC, Olsen DR, Polarek JW;
 PI
 XX WPI; 2004-375406/35.
 DR
 XX N-PSDB; ADN35280.
 DR
 XX Tissue sealant composition comprises cross linking agent and a synthetic
 PT collagen or a synthetic gelatin in a dry state.
 PT
 XX Claim 20; SEQ ID NO 3; 75pp; English.
 PS
 XX The present invention relates to a tissue sealant composition comprises a
 CC crosslinking agent and a synthetic collagen or a synthetic gelatin in a
 CC dry state, they react with each other upon contact with an environment
 CC comprising a physiological pH to form a compound. The compound is useful
 CC to seal a wound comprising a surgical incision (comprising an
 CC angioplasty), laceration or a puncture and contains controlled and
 CC reproducible materials with minimal or no risk of infectivity and disease

CC transmission. The present sequence represents synthetic collagen used as
CC sealant tissue protein.
XX
SQ Sequence 1313 AA;

Query Match 77.4%; Score 48; DB 8; Length 1313;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSGPG 10
||:|||||
Db 1017 GERGGSGSPG 1026

RESULT 54
ADN35277
ID ADN35277 standard; protein; 1313 AA.

XX
AC ADN35277;
XX
DT 01-JUL-2004 (first entry)

XX Helical domain of type III collagen.

XX tissue sealant; synthetic gelatin; Vulnerary; surgical incision;
XX angioplasty; laceration; disease transmission.

XX Homo sapiens.

XX WO2004028404-A2.

XX 08-APR-2004.

XX 30-SEP-2003; 2003WO-US031006.

XX 30-SEP-2002; 2002US-0415309P.

XX (FIBR-) FIBROGEN INC.

XX Yang C, Chang RC, Olsen DR, Polarek JW;

XX WPI; 2004-375406/35.

XX Tissue sealant composition comprises cross linking agent and a synthetic
PT collagen or a synthetic gelatin in a dry state.

XX Claim 20; SEQ ID NO 1; 75pp; English.

XX The present invention relates to a tissue sealant composition comprises a
CC crosslinking agent and a synthetic collagen or a synthetic gelatin in a
CC dry state, they react with each other upon contact with an environment
CC comprising a physiological pH to form a compound. The compound is useful
CC to seal a wound comprising a surgical incision (comprising an
CC angioplasty), laceration or a puncture and contains controlled and
CC reproducible materials with minimal or no risk of infectivity and disease
CC transmission. The present sequence represents the helical domain of type
CC III collagen.

XX Sequence 1313 AA;

Query Match 77.4%; Score 48; DB 8; Length 1313;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSGPG 10
||:|||||
Db 1017 GERGGSGSPG 1026

RESULT 55
AAE02537
ID AAE02537 standard; protein; 1466 AA.

XX

AAE02537;

XX 10-AUG-2001 (first entry)

XX Porcine alpha(III) collagen.

XX Porcine; alpha(III) collagen; gelatin; cytostatic; viral infection;
KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
KW rheumatoid arthritis; beverage; photographic application.

XX Sus scrofa.

XX WO200134647-A2.

XX 17-MAY-2001.

XX 10-NOV-2000; 2000WO-US030792.

XX 12-NOV-1999; 99US-00439058.

XX 10-NOV-2000; 2000US-00709700.

XX (FIBR-) FIBROGEN INC.

XX Bell MP, Neff TB, Polarek JW, Seeley TW;

XX WPI; 2001-335911/35.

XX N-PSDB; AAD06578.

XX Novel isolated and purified bovine or porcine collagens and gelatins
PT useful in medical, pharmaceutical, food and cosmetic industries, as
PT vaccine, and for treating autoimmune disorders, infections and cancer.

XX Example 5; Fig 12; 168pp; English.

XX The present sequence is porcine alpha(III) collagen. The present
CC invention relates to recombinant synthesis of collagens and gelatins
CC derived from animals. Collagen is useful in medical, pharmaceutical, food
CC and cosmetic industries. Collagen is an important component of arterial
CC sealants, bone grafts, drug delivery system, dermal implants, haemostats,
CC and incontinence implants, and for treating autoimmune disorders such as
CC rheumatoid arthritis. Collagen is useful in food products such as sausage
CC casings, and in cosmetics or facial and skin products such as
CC moisturisers. Recombinant gelatin is useful in vaccine formulations for
CC treating viral infections, autoimmune diseases and cancer. Gelatin is
CC useful in the manufacture or as a component of various pharmaceutical and
CC medical devices and products, in food and beverage industries, in hair
CC care and skin care products, as a glue or adhesive in various
CC manufacturing processes, as a light-sensitive coating in various
CC electronic devices, as photoresist base in photolithographic processes,
CC in printing and photographic applications, in laboratory application, and
CC as a component in various gels used for biochemical and electrophoretic
CC analysis, including enzymographic gels

XX Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSGPG 10

Db 1171 GERGGSGSPG 1180

RESULT 56
AAE02534
ID AAE02534 standard; protein; 1466 AA.

XX AAE02534;

XX 10-AUG-2001 (first entry)

XX

```

DE XX Bovine alpha1(III) collagen #2.
KW XX Bovine; alpha1(III) collagen; gelatin; cytostatic; viral infection;
KW XX pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
KW XX medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
KW XX rheumatoid arthritis; beverage; photographic application.
OS XX Bos sp.
XX XX WO200134647-A2.
XX XX 17-MAY-2001.
XX XX 10-NOV-2000; 2000WO-US030792.
XX XX 12-NOV-1999; 99US-00439058.
XX XX 10-NOV-2000; 2000US-00709700.
XX XX (FIBR-) FIBROGEN INC.
XX XX Bell MP, Neff TB, Polarek JW, Seeley TW;
XX XX WPI; 2001-335911/35.
XX XX N-PSDB; AAD06574.
XX XX Novel isolated and purified bovine or porcine collagens and gelatins
XX XX useful in medical, pharmaceutical, food and cosmetic industries, as
XX XX vaccine, and for treating autoimmune disorders, infections and cancer.
XX XX Example 2; Fig 6; 168pp; English.
XX XX The present sequence is bovine alpha1(III) collagen. The present
XX XX invention relates to recombinant synthesis of collagens and gelatins
XX XX derived from animals. Collagen is useful in medical, pharmaceutical, food
XX XX and cosmetic industries. Collagen is an important component of arterial
XX XX sealants, bone grafts, drug delivery system, dermal implants, haemostats,
XX XX and incontinence implants, and for treating autoimmune disorders such as
XX XX rheumatoid arthritis. Collagen is useful in food products such as sausage
XX XX casings, and in cosmetics or facial and skin products such as
XX XX moisturisers. Recombinant gelatin is useful in vaccine formulations for
XX XX treating viral infections, autoimmune diseases and cancer. Gelatin is
XX XX useful in the manufacture or as a component of various pharmaceutical and
XX XX medical devices and products, in food and beverage industries, in hair
XX XX care and skin care products, as a glue or adhesive in various
XX XX manufacturing processes, as a light-sensitive coating in various
XX XX electronic devices, as photoresist base in photolithographic processes,
XX XX in printing and photographic applications, in laboratory application, and
XX XX as a component in various gels used for biochemical and electrophoretic
XX XX analysis, including enzymographic gels
XX XX Sequence 1466 AA;
SQ Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKGAGGSGPG 10
DB 1170 GERGEGSGPG 1179
||:|||||
RESULT 57
AAE02533
ID AAE02533 standard; protein; 1466 AA.
XX XX
XX XX AAE02533;
XX XX 10-AUG-2001 (first entry)
XX XX Bovine alpha1(III) collagen #1.
XX XX Bovine; alpha1(III) collagen; gelatin; cytostatic; viral infection;
XX XX pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
XX XX medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
KW XX

KW Bovine alpha1(III) collagen #2.
OS XX Bos sp.
XX XX WO200134647-A2.
XX XX 17-MAY-2001.
XX XX 10-NOV-2000; 2000WO-US030792.
XX XX 12-NOV-1999; 99US-00439058.
XX XX 10-NOV-2000; 2000US-00709700.
XX XX (FIBR-) FIBROGEN INC.
XX XX Bell MP, Neff TB, Polarek JW, Seeley TW;
XX XX WPI; 2001-335911/35.
XX XX N-PSDB; AAD06574.
XX XX Novel isolated and purified bovine or porcine collagens and gelatins
XX XX useful in medical, pharmaceutical, food and cosmetic industries, as
XX XX vaccine, and for treating autoimmune disorders, infections and cancer.
XX XX Example 2; Fig 6; 168pp; English.
XX XX The present sequence is bovine alpha1(III) collagen. The present
XX XX invention relates to recombinant synthesis of collagens and gelatins
XX XX derived from animals. Collagen is useful in medical, pharmaceutical, food
XX XX and cosmetic industries. Collagen is an important component of arterial
XX XX sealants, bone grafts, drug delivery system, dermal implants, haemostats,
XX XX and incontinence implants, and for treating autoimmune disorders such as
XX XX rheumatoid arthritis. Collagen is useful in food products such as sausage
XX XX casings, and in cosmetics or facial and skin products such as
XX XX moisturisers. Recombinant gelatin is useful in vaccine formulations for
XX XX treating viral infections, autoimmune diseases and cancer. Gelatin is
XX XX useful in the manufacture or as a component of various pharmaceutical and
XX XX medical devices and products, in food and beverage industries, in hair
XX XX care and skin care products, as a glue or adhesive in various
XX XX manufacturing processes, as a light-sensitive coating in various
XX XX electronic devices, as photoresist base in photolithographic processes,
XX XX in printing and photographic applications, in laboratory application, and
XX XX as a component in various gels used for biochemical and electrophoretic
XX XX analysis, including enzymographic gels
XX XX Sequence 1466 AA;
SQ Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GKGAGGSGPG 10
DB 1170 GERGEGSGPG 1179
||:|||||
RESULT 58
ABB50291
ID ABB50291 standard; protein; 1466 AA.
XX XX
XX XX ABB50291;
XX XX 08-FEB-2002 (first entry)
XX XX Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.
XX XX Ovarian tumour marker gene; human; overexpression; upregulation;
XX XX epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;
XX XX identification; serous cystadenoma; borderline serous tumour;
XX XX serous cystadenocarcinoma; mucinous cystadenocarcinoma;
XX XX mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;
XX XX undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;
XX XX adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;

```

KW immune response pathway; cell proliferation regulation; protein folding;
KW membrane localised; secreted; therapeutic target; cytostatic;
KW gene therapy; vaccine.

XX Homo sapiens.

XX WO200175177-A2.

XX 11-OCT-2001.

XX 03-APR-2001; 2001WO-US010947.

XX 03-APR-2000; 2000US-0194336P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;

XX WPI; 2001-626450/72.

XX N-PSDB; ABA83117.

XX Detecting and identifying ovarian tumor, identifying increased risk for
XX developing ovarian cancer, and determining effectiveness of ovarian
XX cancer treatment, by measuring expression level of ovarian tumor marker
XX gene.

XX Claim 23; Page 114-117; 140pp; English.

XX The invention relates to methods for diagnosing and prognosing ovarian
XX tumours in an individual via the detection and measurement of the
XX expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,
XX ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,
XX ABA83181 and ABA83183). The methods of the invention are useful for
XX detecting an ovarian tumour in a patient, for identifying an individual
XX at increased risk for developing ovarian cancer, in prognostic tests for
XX assessing the relative severity of ovarian cancer, in tests for
XX monitoring a patient in remission from ovarian cancer and in tests for
XX monitoring disease status in a patient being treated for ovarian cancer.
XX The methods can additionally be used to identify a particular tumour as
XX being an ovarian tumour (i.e., an epithelial ovarian tumour selected from
XX serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,
XX mucinous cystadenoma, borderline mucinous tumour, mucinous
XX cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,
XX clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner
XX tumour. The ovarian tumour marker genes of the invention were identified
XX using SAGE (serial analysis of gene expression) and were found to be
XX overexpressed in a broad variety of ovarian epithelial tumour cells
XX relative to normal ovarian epithelial cells. The marker genes are
XX implicated in immune response pathways, in the regulation of cell
XX proliferation and in protein folding, and many of these are membrane-
XX localised or secreted. In addition to their use as diagnostic and
XX prognostic markers, the ovarian tumour marker genes or their encoded
XX proteins may be used as therapeutic targets for the treatment and
XX prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent
XX proteins encoded by ovarian tumour marker genes of the invention

XX Sequence 1466 AA;

XX Query Match 77.4%; Score 48; DB 4; Length 1466;
XX Best Local Similarity 80.0%; Pred. No. 69;
XX Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10

|||:|||||

Db 1170 GERSEGSPPG 1179

RESULT 59

ABB90747

ID ABB90747 standard; protein; 1466 AA.

XX AC ABB90747;

XX

DT 30-MAY-2002 (first entry)

XX Human Tumour Endothelial Marker polypeptide SEQ ID NO 226.

XX Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic;
XX normal endothelial marker; pan-endothelial marker; immunostimulant;
XX antiangiogenic; tumour; neoangiogenesis; vascularised tumour;
XX polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;
XX psoriasis.

XX Homo sapiens.

XX WO200210217-A2.

XX 07-FEB-2002.

XX 01-AUG-2001; 2001WO-US024031.

XX 02-AUG-2000; 2000US-022599P.

XX 11-AUG-2000; 2000US-0224360P.

XX 11-APR-2001; 2001US-0282850P.

XX (UYJO) UNIV JOHNS HOPKINS.

XX St Croix B, Kinzler KW, Vogelstein B;

XX WPI; 2002-291856/33.

XX N-PSDB; ABL92101.

XX An isolated molecule comprising an antibody variable region which
XX specifically binds to an extracellular domain of a tumor endothelial
XX marker (TEM) protein, useful for inhibiting tumor growth.

XX Claim 68; Page 197-200; 331pp; English.

XX The invention relates to an isolated molecule comprising an antibody
XX variable region which specifically binds to an extracellular domain of a
XX tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,
XX ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
XX proteins have cytostatic, immunostimulant and antiangiogenic activity.
XX They are useful for inhibiting tumour growth, neoangiogenesis in subjects
XX bearing a vascularised tumour, polycystic kidney disease, diabetic
XX retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM
XX genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)
XX are disclosed, as are marker oligonucleotide sequences: tumour
XX endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal
XX endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers
XX (PEM) ABL91903-ABL91995

XX Sequence 1466 AA;

XX Query Match 77.4%; Score 48; DB 5; Length 1466;
XX Best Local Similarity 80.0%; Pred. No. 69;
XX Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10

|||:|||||

Db 1170 GERSEGSPPG 1179

RESULT 60

ABUS4454

ID ABUS4454 standard; protein; 1466 AA.

XX AC ABUS4454;

DT 12-MAR-2003 (first entry)

XX Human tumour endothelial marker TEM 15.

XX Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;

XX Tumour endothelial marker; normal endothelial marker; PEM;

XX pan-endothelial marker; polycystic kidney disease; psoriasis;

KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;
KW neangiogenesis; immune response; cytostatic; antidiabetic;
KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic.
XX
XX Homo sapiens.
XX WO200283874-A2.
XX
XX 24-OCT-2002.
XX
XX 10-APR-2002; 2002WO-US008253.
XX
XX 11-APR-2001; 2001US-0282850P.
XX 06-FEB-2002; 2002US-0354262P.
XX
XX (UYJO) UNIV JOHNS HOPKINS.
XX
XX Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;
XX
XX WPI; 2003-093016/08.
XX N-PSDB; ABX72026.
XX
XX New purified human transmembrane protein, designated as tumor endothelial
XX marker (TEM) 3, useful for detecting, diagnosing or treating tumors,
XX polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or
XX psoriasis.
XX
XX Disclosure; Page 211-214; 374pp; English.
XX
XX The present invention relates to a novel method for the isolation of
XX endothelial cells (ECs), and the identification of genes expressed in
XX normal and tumor ECs. Tumour endothelial marker (TEM), normal
XX endothelial marker (NEM), and pan-endothelial marker (PEM) genes are
XX identified in human ECs. The human EC marker proteins and the
XX polynucleotide sequences encoding them are useful for detecting,
XX diagnosing or treating tumors as well as polycystic kidney disease,
XX diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also
XX useful for inhibiting neangiogenesis or tumour angiogenesis, for
XX inducing an immune response to tumour endothelial cells in a patient, or
XX for identifying candidate drugs for treating tumours. The present
XX sequence represents a human TEM or NEM protein of the invention
XX
XX Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 6; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGPG 10
||:|:|||||
DB 1170 GERGEGSGPG 1179

RESULT 61
ABR47418
ID ABR47418 standard; protein; 1466 AA.
XX
XX ABR47418;
XX
XX 12-JUN-2003 (first entry)
XX
XX Breast cancer associated protein sequence SEQ ID NO:68.
XX
XX Human; breast cancer; cytostatic; gene therapy.
XX
XX Homo sapiens.
XX
XX WO2003004989-A2.
XX
XX 16-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-US019669.
XX

PR 21-JUN-2001; 2001US-0299887P.
PR 27-JUN-2001; 2001US-0301572P.
PR 18-JUL-2001; 2001US-0306501P.
PR 25-SEP-2001; 2001US-0325002P.
PR 05-MAR-2002; 2002US-0362585P.
PR 14-MAY-2002; 2002US-0380391P.
XX
XX (MILL-) MILLENIUM PHARM INC.
XX
XX Lillie J, Gannavathu M, Glatt K, Hoersh S, Kamatkar S;
XX Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
XX Bast RC, Hortobagyi GN, Puzstai L, Meric F, Sahin A, Mills GB;
XX
XX WPI; 2003-210381/20.
XX N-PSDB; ACC50110.
XX
XX Breast cancer diagnosis or treatment by comparing the level of expression
XX of a marker in a patient sample with that in the control non-breast
XX cancer sample.
XX
XX Claim 1; SEQ ID NO 68; 128pp; English.
XX
XX The present invention describes a method for assessing whether a patient
XX is afflicted with breast cancer. The method comprises comparing the level
XX of expression of a marker (Gene/polypeptide see ACC50076 to ACC50334 and
XX ABR47386 to ABR47632) in a patient sample and the normal level of
XX expression of the marker in a control non-breast cancer sample, where a
XX significant increase in the level of expression of the marker in the
XX patient sample and the normal level is an indication that the patient is
XX afflicted with breast cancer. The breast cancer associated sequences from
XX the present invention have cytostatic activities and can be used in gene
XX therapy. The method is useful for diagnosing and treating breast cancer.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 6; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGGAGSGPG 10
||:|:|||||
DB 1170 GERGEGSGPG 1179

RESULT 62
ADP65248
ID ADP65248 standard; protein; 1466 AA.
XX
XX ADP65248;
XX
XX 12-AUG-2004 (first entry)
XX
XX Human alpha 1 type III collagen preproprotein.
XX
XX autoimmune disease; arthritis; gene expression analysis;
XX rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;
XX antiarthritic; osteopathic; antigout; antiinflammatory; dermatological;
XX immunomodulatory; lupus; ankylosing spondylitis; Fibrositis;
XX fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;
XX immune; human.
XX
XX Homo sapiens.
XX
XX WO2003072827-A1.
XX
XX 04-SEP-2003.
XX
XX 31-OCT-2002; 2002WO-US035433.
XX
XX 31-OCT-2001; 2001US-0336220P.
XX

KW Alzheimer's disease; vascular-related dementia;
KW Creutzfeldt-Jacob disease; bovine spongiform encephalopathy;
KW Parkinson's disease; brain trauma; multiple sclerosis;
KW peripheral disorder; septic shock; hepatic encephalopathy;
KW diabetic hypertension; diabetic microangiopathy; sleeping sickness;
KW neuropsychiatric disorder; depression; autism; schizophrenia; psychosis;
KW CNS disorder; brain tumour; epilepsy; migraine; narcolepsy; insomnia;
KW chronic fatigue syndrome; mountain sickness; AIDS-related dementia;
KW angiogenesis-related disorder; proliferative vitreoretinopathy;
KW rheumatoid arthritis; Crohn's disease; atherosclerosis;
KW ovarian hyperstimulation; psoriasis; endometriosis; neovascularisation;
KW restenosis; balloon angioplasty; scar tissue overproduction;
KW peripheral vascular disease; hypertension; inflammatory vasculitides;
KW Reynaud's disease; Reynaud's phenomenon; aneurysm; arterial restenosis;
KW thrombophlebitis; lymphangitis; lymphedema; wound healing; tissue repair;
KW ischaemia reperfusion injury; angina; myocardial infarction;
KW chronic heart condition; osteoporosis; type III, alpha 1 collagen;
COL3A1.
XX
XX Homo sapiens.
XX WO2004056386-A2.
XX PD 08-JUL-2004.
XX PF 19-DEC-2003; 2003WO-NL000915.
XX PR 19-DEC-2002; 2002EP-00080503.
XX PR 25-APR-2003; 2003US-0465234P.
XX PA (UYLE-) RIJKSUNIV LEIDEN.
XX PI Gaillard PJ, De Boer AG, Brink A;
XX WPI; 2004-500264/47.
XX
XX Modulating the permeability of endothelial cells, useful for treating or
XX preventing e.g. neurodegenerative disorders, comprises altering in the
XX endothelial cells the activity or the steady-state level of a pro-barrier
XX polypeptide.
XX
XX Claim 1; SEQ ID NO 45; 222pp; English.
XX
XX This sequence represents type III, alpha 1 collagen encoded by the COL3A1
XX gene. This protein is upregulated in BCEC (brain capillary endothelial
XX cells) by physical co-culture with astrocytes. This protein sequence may
XX be used in the method of the invention for modulating the permeability of
XX endothelial cells. The method comprises altering in the endothelial cells
XX the activity or the steady-state level of a pro-barrier (PB) polypeptide
XX having an amino acid sequence having at least 90% identity with a claimed
XX amino acid sequence selected from ADQ26047ADQ26048ADQ26049ADQ26050ADQ26051
XX 1ADQ26052ADQ26053ADQ26054ADQ26055ADQ26056ADQ26057ADQ26058ADQ26059ADQ26060
XX ADQ26061ADQ26062ADQ26063ADQ26064ADQ26065ADQ26066ADQ26067ADQ26068ADQ26069A
XX ADQ26070ADQ26071ADQ26072ADQ26073ADQ26074ADQ26075ADQ26076ADQ26077ADQ26078AD
XX Q26079ADQ26080ADQ26081ADQ26082ADQ26083ADQ26084ADQ26085ADQ26086ADQ26087ADQ
XX 26088ADQ26089ADQ26090ADQ26091ADQ26092ADQ26093ADQ26094ADQ26095ADQ26096ADQ2
XX 6097ADQ26098ADQ26099. The PB polypeptides, nucleic acid molecules
XX encoding them, antagonists or a gene therapy vector comprising an
XX antisense nucleotide sequence capable of inhibiting the expression of the
XX nucleotide sequence encoding a PB polypeptide, are useful in the
XX manufacture of a composition for treating or preventing a microvascular
XX permeability-modifying disorder, including neurodegenerative disorders
XX (e.g. cerebrovascular accidents, Alzheimer's disease, vascular-related
XX dementia, Creutzfeldt-Jacob disease, bovine spongiform encephalopathy,
XX Parkinson's disease, brain trauma, multiple sclerosis), peripheral
XX disorders with a CNS component (such as septic shock, hepatic
XX encephalopathy, (diabetic) hypertension, diabetic microangiopathy or
XX sleeping sickness), neuropsychiatric disorders (e.g. depression, autism,
XX schizophrenia and other psychoses; other CNS disorders (brain tumours,
XX epilepsy, migraine, narcolepsy, insomnia, chronic fatigue syndrome,
XX mountain sickness, or AIDS-related dementia); and angiogenesis-related
XX disorders (proliferative vitreoretinopathy, rheumatoid arthritis, Crohn's
XX disease, atherosclerosis, ovarian hyperstimulation, psoriasis,

CC endometriosis associated with neovascularisation, restenosis subsequent
CC to balloon angioplasty, scar tissue overproduction, peripheral vascular
CC disease, hypertension, inflammatory vasculitides, Reynaud's disease,
CC Reynaud's phenomenon, aneurysm, arterial restenosis, thrombophlebitis,
CC lymphangitis, lymphedema, wound healing and tissue repair, ischaemia
CC reperfusion injury, angina, myocardial infarctions, chronic heart
CC conditions, or osteoporosis). The PB polypeptide may also be used in the
CC manufacture of a composition for reversibly increasing the microvascular
CC permeability in a subject.
XX
XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 8; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGAGGSPG 10
||:|:||||
DB 1170 GERGEGSPG 1179

RESULT 65
ADQ29677
ID ADQ29677 standard; protein; 1466 AA.
XX
XX AC ADQ29677;

XX DT 07-OCT-2004 (first entry)

XX DE Human colorectal cancer-associated protein #32.

XX KW human; colon cancer; TIMP1; Regl-alpha;
XX colorectal cancer-associated marker.

XX OS Homo sapiens.

XX PN EP1439393-A2.

XX PD 21-JUL-2004.

XX PF 15-DEC-2003; 2003EP-00257868.

XX PR 13-DEC-2002; 2002US-0433554P.

XX PR 31-JUL-2003; 2003US-0491397P.

XX PA (FARB) BAYER HEALTHCARE LLC.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Astle JH, Boardman LA, Bugart LJ, Burgess CC, Catino TJ;
XX Dwivedi P, Huntress M, Johnson KA, Lewis ME, Maimonis PJ,
XX Brown-Shimer SLA, Thiagalingam A, Thibodeau SN, Molino GA;

XX WPI; 2004-545561/53.

XX DR N-PSDB; ADQ29610.

XX PT Diagnosing colon cancer in individual, preferably human, by detecting
XX presence of TIMP 1 in sample, where presence of TIMP 1 in sample is
XX indicative of colon cancer in individual.

XX PS Claim 7; SEQ ID NO 103; 433pp; English.

XX The invention comprises a method for diagnosing colon cancer in an
XX individual, the method involves obtaining a serum sample from the
XX individual and detecting the presence of either TIMP1 or Regl-alpha and
XX an additional colorectal cancer-associated marker. The method of the
XX invention is useful for diagnosing colon cancer in an individual. The
XX present amino acid sequence represents a human colorectal cancer-
XX associated protein of the invention.

XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 8; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;

```
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGAGGSPG 10
Db 1170 GERSEGGSPG 1179
||:|||||

RESULT 66
ADRI6802
ID ADR16802 standard; protein; 1466 AA.
XX AC ADR16802;
XX DT 21-OCT-2004 (first entry)
XX DE Human collagen III alpha2 (III) fragment protein.
XX KW T-lymphocyte; AAA; abdominal aortic aneurysm; human; collagen III;
XX KW therapy.
XX OS Homo sapiens.
XX PN US2004151731-A1.
XX PD 05-AUG-2004.
XX PF 04-FEB-2003; 2003US-00357851.
XX PR 04-FEB-2003; 2003US-00357851.
XX PA (JICH/) JICHA D L.
XX PI Jicha DL;
XX DR WPI; 2004-561474/54.
XX PT New isolated and purified T-lymphocyte derived from abdominal aortic
XX PT tissues useful to prevent/treat an abdominal aortic aneurysm or rupture.
XX PS Disclosure; SEQ ID NO 3; 24pp; English.
XX CC The present invention relates to a T-lymphocyte derived from abdominal
XX CC aortic tissue which are specifically reactive with collagen I, collagen
XX CC III or their fragments. The invention is useful to prevent or treat an
XX CC abdominal aortic aneurysm (AAA) or rupture in a mammal. The present
XX CC sequence is human collagen I alpha2 (III) fragment protein. This sequence
XX CC is used in the invention.
XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 8; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGAGGSPG 10
Db 1170 GERSEGGSPG 1179
||:|||||

RESULT 67
ADRI6427
ID ADR16427 standard; protein; 1466 AA.
XX AC ADR16427;
XX DT 21-OCT-2004 (first entry)
XX DE Human collagen I alpha2 (III) fragment protein.
XX KW T-lymphocyte; AAA; abdominal aortic aneurysm; vaccine; therapy; human;
XX KW collagen I.
XX OS Homo sapiens.
```

```
XX US2004151732-A1.
XX PN 05-AUG-2004.
XX PD 04-FEB-2003; 2003US-00358024.
XX PF 04-FEB-2003; 2003US-00358024.
XX PR (JICH/) JICHA D L.
XX PA (PELU/) PELUSE S.
XX PI Jicha DL, Peluse S;
XX DR WPI; 2004-570707/55.
XX DE Isolated or purified lymphocytes derived from blood in abdominal aortic
XX DE aneurysm patients useful for early diagnosis of aneurysms are reactive
XX DE with collagen I, collagen III and/or their fragments.
XX PS Disclosure; SEQ ID NO 3; 22pp; English.
XX CC The present provides a T-lymphocyte derived from blood in abdominal
XX CC aortic aneurysm (AAA) patients which are reactive with collagen I,
XX CC collagen III and their fragments. The invention is useful in diagnosis,
XX CC prevention and treatment of abdominal aortic aneurysm or rupture in a
XX CC mammal. The invention is also useful in the vaccine preparation. The
XX CC present sequence is human collagen I alpha2 (III) fragment protein.
XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 8; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGAGGSPG 10
Db 1170 GERSEGGSPG 1179
||:|||||

RESULT 68
ABM80366
ID ABM80366 standard; protein; 1466 AA.
XX AC ABM80366;
XX DT 18-NOV-2004 (first entry)
XX DE Tumour-associated antigenic target (TAT) polypeptide PRO2665, SEQ:922.
XX KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
XX KW tumour; diagnosis; cell proliferative disorder; breast cancer;
XX KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
XX KW central nervous system cancer; bladder cancer; pancreatic cancer;
XX KW cervical cancer; melanoma; leukaemia; hybridisation probe;
XX KW chromosome identification; chromosome mapping; gene mapping;
XX KW gene therapy; cytostatic.
XX OS Homo sapiens.
XX PN WO2004030615-A2.
XX PD 15-APR-2004.
XX PF 29-SEP-2003; 2003WO-US028547.
XX PR 02-OCT-2002; 2002US-0414971P.
XX PA (GETH ) GENENTECH INC.
XX KW Wu TD, Zhang Z, Zhou Y;
XX KW WPI; 2004-347921/32.
XX DR
```

DR N-PSDB; ACN37799.
 XX New tumor-associated antigenic target polypeptides and nucleic acids, (TAT)
 PT useful in preparing a medicament for treating or detecting a
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
 PT prostate cancer or tumor.
 XX
 PS Claim 12; SEQ ID NO 922; 7273pp; English.
 XX
 CC The invention relates to human tumour-associated antigenic target (TAT)
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are
 CC overexpressed in cancer tissues compared to normal tissues, and may thus
 CC serve as effective targets for the diagnosis and treatment of cancer in
 CC mammals. The invention also relates to nucleic acid and polypeptide
 CC sequences at least 80% identical to the TAT nucleic acids and
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT polypeptide of the invention
 XX
 SQ Sequence 1466 AA;
 Query Match 77.4%; Score 48; DB 8; Length 1466;
 Best Local Similarity 80.0%; Pred. NO. 69;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GKGAGSGSPG 10
 ||:|:|:|:|:|
 Db 1170 GERGSESGSPG 1179
 RESULT 69
 ADR67267
 ID ADR67267 standard; protein; 1466 AA.
 AC ADR67267;
 XX 02-DEC-2004 (first entry)
 XX Human bladder cancer associated amino acid sequence.
 DE bladder cancer tissue; bladder cancer; cytostatic.
 KW Homo sapiens.
 OS WO2004076613-A2.
 XX 10-SEP-2004.
 XX 24-FEB-2004; 2004WO-DE000364.
 PF 26-FEB-2003; 2003DE-01009729.
 XX (HERR/) HERR A.
 PA (HINZ/) HINZMANN B.
 PA (DAHL/) DAHL E.
 PA (STAUB/) STAUB E.
 PA (PILARSKY C.)
 PA (SPEC/) SPECHT T.
 XX Herr A, Hinzmann B, Dahl E, Staub E, Pilarsky C, Specht T;
 XX WPI; 2004-653385/63.
 DR

XX New nucleic acids, and encoded proteins, from bladder cancer tissue,
 PT useful for diagnosis, treatment and in screening for specific binding
 PT agents.
 XX Claim 2; Fig 3; 112pp; German.
 PS
 XX The present invention describes nucleic acids (I) associated with bladder
 CC cancer tissue. Also described: (1) peptides and proteins (II) containing
 CC an amino acid sequence encoded by (1); (2) a method for diagnosing
 CC bladder cancer (BC), or monitoring its progression, that uses (1), (II)
 CC or agents (Z) that inhibit, or bind to, (I) or (II), labelled with a
 CC reporter; and (3) a method for treating BC that uses (I), (II) or (Z).
 CC (I) and (II) have cytostatic activity. (I) and (II) can be used to detect
 CC and monitor progression of) bladder cancer (BC), or the risk of
 CC developing it; to screen for specific binding agents (Z), and to treat
 CC BC. (Z) are also useful as diagnostic and therapeutic agents. The present
 CC sequence represents a human amino acid sequence associated with bladder
 CC cancer, which is used in the exemplification of the present invention.
 XX
 SQ Sequence 1466 AA;
 Query Match 77.4%; Score 48; DB 8; Length 1466;
 Best Local Similarity 80.0%; Pred. NO. 59;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GKGAGSGSPG 10
 ||:|:|:|:|:|
 Db 1170 GERGSESGSPG 1179
 RESULT 70
 ADV70235
 ID ADV70235 standard; protein; 1466 AA.
 XX ADV70235;
 AC 10-MAR-2005 (first entry)
 XX Tumor-associated antigenic target polypeptide TAT474.
 DE cytostatic; diagnosis; therapy; tumor;
 KW tumor-associated antigenic target polypeptide; TAT.
 OS Homo sapiens.
 XX WO2004112829-A2.
 XX 29-DEC-2004.
 XX 21-MAY-2004; 2004WO-US016121.
 XX 23-MAY-2003; 2003US-0473238P.
 PR 27-FEB-2004; 2004US-0548299P.
 XX (GETH) GENENTECH INC.
 XX Phillips H;
 XX WPI; 2005-048766/05.
 DR N-PSDB; ADV70141.
 XX Treating a mammal having a tumor of glial origin comprising cells that
 PT express a type A or B glial tumor antigen by contacting the cells with a
 PT composition comprising first and second binding agents.
 XX Disclosure; SEQ ID NO 159; 374pp; English.
 PS The invention describes a method of treating a mammal having a tumor of
 CC glial origin comprising cells that express a type A or B glial tumor
 CC antigen comprises contacting the cells with a composition of: a first
 CC binding agent comprising a first antibody, oligopeptide or organic
 CC molecule that binds to a type A or B glial tumor antigen; and a second

CC binding agent comprising a second antibody, oligopeptide or organic
CC molecule that binds to a type B or A glial tumor antigen. Also described
CC is a method of determining the presence of a type A or B glial tumor in a
CC mammal. The method is useful in treating a mammal having a tumor of glial
CC origin comprising cells that express a type A or B glial tumor antigen.
CC This sequence represents a human tumor-associated antigenic target
CC polypeptide.
XX
SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 9; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
||:|||||
Db 1170 GERSESGPG 1179

RESULT 71
ADZ09873
ID ADZ09873 standard; protein; 1466 AA.

XX AC ADZ09873;

XX DT 30-JUN-2005 (first entry)

XX DE Human breast cancer marker COL3A1 protein.

XX KW neoplasm; tumor marker; cytostatic; breast tumor; gene therapy; COL3A1.

XX OS Homo sapiens.

XX PN EP1522594-A2.

XX PD 13-APR-2005.

XX PF 30-JUN-2004; 2004EP-00015374.

XX PR 06-OCT-2003; 2003EP-00022587.

XX PA (FARB) BAYER HEALTHCARE AG.

XX PI Munnes M, Bojar H;

XX DR WPI; 2005-265481/28.

XX DR N-PSDB; ADZ09708.

XX PT Characterizing the state of a neoplastic disease in a subject by
PT comparing the pattern of determined expression levels of marker genes in
PT a biological sample with reference patterns of expression levels.

XX PS Disclosure; SEQ ID NO 317; 542pp; English.

XX CC The invention relates to a novel method for characterizing the state of a
CC neoplastic disease in a subject comprising determining the pattern of
CC expression levels of at least 6, 8, 10, 15, 20, 30, 47 or 67 marker genes
CC in a biological sample, comparing the pattern of expression levels with
CC reference patterns of expression levels and characterizing the state of
CC the neoplastic disease in the subject from the outcome of the comparison.
CC The method of the invention demonstrates cytostatic applications and may
CC be useful in characterizing the state of a neoplastic disease in a
CC subject, preferably breast cancer, in gene therapy and for screening for
CC compounds having a curative effect on a neoplastic disease. The current
CC sequence is that of the human breast cancer marker COL3A1 protein of the
CC invention which is differentially expressed in neoplastic tissue.

XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 9; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
||:|||||
Db 1170 GERSESGPG 1179

RESULT 72
AEA04495

ID AEA04495 standard; protein; 1466 AA.

XX AC AEA04495;

XX DT 28-JUL-2005 (first entry)

XX DE Human protein from gene overexpressed in cancer, COL3A1.

XX KW Tumor marker; colon tumor; cancer; cytostatic; neoplasm; diagnostic;
XX KW microarray; drug screening.

XX OS Homo sapiens.

XX PN WO2005044990-A2.

XX PD 19-MAY-2005.

XX PF 01-NOV-2004; 2004WO-US036404.

XX PR 04-NOV-2003; 2003US-00700439.

XX PA (FARB) BAYER HEALTHCARE LLC.

XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX PI Burgess C, Myerow S, Thiagalingam A, Maimonis P, Molino G;

XX PI Burgart L, Boardman LA, Thibodeau S, Lewis M;

XX DR WPI; 2005-372198/38.

XX DR N-PSDB; AEA04402.

XX DR REFSEQ; NP_000081.

XX PT Detecting expression of one or more nucleic acid sequences in biological
PT sample, useful for detecting cancer, comprises detecting a change in the
PT expression level of one or more nucleic acid sequences relative to a
PT control expression level.

XX PS Claim 20; SEQ ID NO 140; 256pp; English.

XX CC The invention relates to detecting differential expression of one or more
CC nucleic acid sequences (appearing as AEA04356-AEA04448 in a biological
CC sample comprising obtaining the sample from a subject, and detecting a
CC change in the expression level of one or more nucleic acid sequences
CC relative to a control expression level of the nucleic acid sequences, is
CC new. Also included are detecting cancer (or a pre-malignant condition
CC thereof) in a subject (comprising comparing the expression level of one
CC or more nucleic acid sequences in a biological sample from the subject
CC with a control expression level of the nucleic acid sequences, where a
CC change of at least two-fold in the expression level of the nucleic acid
CC sequences is indicative of cancer or pre-malignant condition), monitoring
CC the onset (or progression, or regression) of cancer (or a pre-malignant
CC condition) in a subject (by detecting in a biological sample of the
CC subject at a first point in time the expression of one or more nucleic
CC acid sequences, repeating the first step at a subsequent point in time
CC and comparing the expression level detected, where a change in the
CC expression level is indicative of progression of cancer or its pre-
CC malignant condition in the subject), determining prognosis for cancer or
CC its pre-malignant condition in a subject (comprising detecting in a
CC biological sample of the subject, the expression level of one or more
CC nucleic acid sequences, comparing the expression level with a reference
CC expression level of the nucleic acid sequences and evaluating the
CC prognosis of the subject based on the comparison), determining the
CC efficacy of a test compound for inhibiting cancer in a subject,
CC determining the efficacy of a therapy for inhibiting cancer in a subject,
CC selecting a composition for inhibiting cancer in a subject, inhibiting
CC cancer in a subject, a polypeptide encoded by the nucleic acids above
CC (appearing as AEA04449-AEA04541), an antibody that specifically binds to

CC the polypeptide sequence, and detecting in a biological sample the
CC presence of a polypeptide. The method is useful for detecting
CC differential expression of one or more nucleic acid sequences in a
CC biological sample, which is useful for detecting cancer (especially colon
CC cancer), monitoring the onset, progression, or regression of cancer or a
CC pre-malignant condition, or determining prognosis for cancer or its pre-
CC malignant condition in a subject, or for determining the efficacy of a
CC test compound for inhibiting cancer in a subject. The compound is useful
CC for inhibiting cancer in a subject. The antibodies may also be used to
CC treat cancer. The present sequence is a protein from a human gene over-
CC expressed in cancer samples.

XX SQ Sequence 1466 AA;

Query Match 77.4%; Score 48; DB 9; Length 1466;
Best Local Similarity 80.0%; Pred. No. 69;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEKGAGSGSPG 10

Db 1170 GERSEGSFG 1179
|||:|||||

RESULT 73

ABG15191
ID ABG15191 standard; protein; 1469 AA.

XX AC ABG15191;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #15182.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US008631.

XX PR 31-MAR-2000; 2000US-00540217.

XX PR 23-AUG-2000; 2000US-00649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS79378.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensic, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 45550; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1469 AA;

Query Match 77.4%; Score 48; DB 4; Length 1469;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEKGAGSGSPG 10

Db 1173 GERSEGSFG 1182
|||:|||||

RESULT 74

ADE09399
ID ADE09399 standard; protein; 1470 AA.

XX AC ADE09399;

DT 29-JAN-2004 (first entry)

XX DE Novel protein-related contig polypeptide sequence #465.

XX novel gene; novel protein; tissue marker; molecular weight marker;
KW chromosome marker; genetic disorder; contig.

XX OS Unidentified.

XX PN WO2003054152-A2.

XX PD 03-JUL-2003.

XX PF 10-DEC-2002; 2002WO-US039555.

XX PR 10-DEC-2001; 2001US-0339739P.

XX PR 11-DEC-2001; 2001US-0339453P.

XX PR 14-MAR-2002; 2002US-0365091P.

XX PR 14-MAR-2002; 2002US-0365384P.

XX PR 12-APR-2002; 2002US-0372381P.

XX PR 12-APR-2002; 2002US-0372615P.

XX PR 22-APR-2002; 2002US-00128558.

XX PR 24-APR-2002; 2002US-0376045P.

XX PA (HYSE-) HYSEQ INC.

XX Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;

PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;

PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;

XX WPI; 2003-569235/53.

XX New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.

XX Disclosure; SEQ ID NO 2943; 1177pp; English.

XX The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence was used in the
CC exemplification of the invention.

```
XX
SQ Sequence 1470 AA;
Query Match 77.4%; Score 48; DB 7; Length 1470;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGKAEKSGPG 10
Db 1174 GERSEKSGPG 1183

RESULT 75
ADU04510
ID ADU04510 standard; protein; 1572 AA.
XX
AC ADU04510;
XX
DT 13-JAN-2005 (first entry)
XX
DE SLPI-Collagen III amino acid sequence, seq id 27.
XX
KW Antiinflammatory; vulnerary; pro-alpha; laminin glycoprotein;
KW secretory leukocyte protease inhibitor; wound; fibrotic disorder.
XX
OS Unidentified.
XX
FN GB2400852-A.
XX
PD 27-OCT-2004.
XX
PF 21-OCT-2003; 2003GB-00024457.
XX
PR 22-APR-2003; 2003GB-00009064.
XX
PA (UYMA-) UNIV VICTORIA MANCHESTER.
XX
PI Kadler K, Bulleid N, Ashcroft G;
XX
DR WPI; 2004-768581/76.
XX
XX Novel modified pro-alpha chain, useful for treating wound or fibrotic
PT disorder, having triple helical forming domain linked to N-terminal
PT domain having polypeptide of laminin glycoprotein or secretory leukocyte
PT protease inhibitor.
XX
PS Claim 18; SEQ ID NO 27; 59pp; English.
XX
XX The invention relates to a modified pro-alpha chain (I), comprising a
CC triple helical forming domain linked to at least one N-terminal domain
CC having polypeptide sequence from at least part of a laminin glycoprotein,
CC or at least part of a secretory leukocyte protease inhibitor or its
CC functional derivatives. Further disclosed is a DNA molecule (II) encoding
CC (I), a procollagen molecule (III) comprising a trimer of pro-alpha
CC chains, where the pro-alpha chain is (I). Compositions of the invention
CC are useful in the treatment of medical conditions such as wounds or
CC fibrotic disorders. Even if (I) has very short half-life, it can be
CC continuously administered for treating wounds. The current sequence
CC represents the SLPI-Collagen III amino acid sequence.
XX
SQ Sequence 1572 AA;
Query Match 77.4%; Score 48; DB 8; Length 1572;
Best Local Similarity 80.0%; Pred. No. 75;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGKAEKSGPG 10
Db 1276 GERSEKSGPG 1285
```

Search completed: March 11, 2006, 12:00:51
Job time : 331.8 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:01:18 ; Search time 28.2 Seconds
(without alignments)
40.943 Million cell updates/sec

Title: US-10-698-121A-2
Perfect score: 62
Sequence: 1 GKGAGSPGILL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	93.5	633	2 B40983	collagen alpha 1(X)
2	52	83.9	1603	2 S23810	collagen alpha 1(X)
3	50	80.6	618	2 S32436	collagen alpha 2(I)
4	49	79.0	650	2 T22002	hypothetical prote
5	49	79.0	680	2 S31216	collagen alpha 1(X)
6	49	79.0	1712	1 CGHU2B	collagen alpha 2(I)
7	48	77.4	636	2 S41067	collagen alpha 1(I)
8	48	77.4	1049	1 CGB07S	collagen alpha 1(I)
9	48	77.4	1464	2 S53856	collagen alpha 1(I)
10	48	77.4	1466	1 CGHU7L	collagen alpha 1(I)
11	47	75.8	244	2 JCA708	gelatin-binding 28
12	47	75.8	1027	2 S28774	collagen alpha cha
13	47	75.8	1349	2 I48103	type VII collagen
14	47	75.8	1691	1 S22917	collagen alpha 5(I)
15	46	74.2	310	1 I50696	collagen alpha 1(I)
16	46	74.2	754	2 A55267	collagen alpha 5(I)
17	46	74.2	812	2 S31521	collagen COLFI - f
18	46	74.2	886	2 I50694	collagen alpha 1(I)
19	46	74.2	888	2 S28791	collagen alpha 1(X)
20	46	74.2	920	2 A45748	collagen alpha 1(V)
21	46	74.2	1024	2 T45467	collagen alpha 1(X)
22	46	74.2	1418	2 T45467	collagen alpha 1(I)
23	46	74.2	1419	1 CGH182	collagen alpha 1(I)
24	46	74.2	1487	1 CGHU6C	collagen alpha 1(I)
25	46	74.2	1487	2 B41182	collagen alpha 1(I)
26	46	74.2	1492	2 A40333	collagen alpha 1(X)
27	46	74.2	1546	1 CGHU2E	collagen alpha 2(X)
28	46	74.2	1669	1 CGWS4B	collagen alpha 1(I)
29	46	74.2	1744	2 S40991	collagen alpha 1(I)

30	46	74.2	1806	1 CGHU1E	collagen alpha 1(X)
31	45	72.6	198	2 I49558	collagen alpha 1(I)
32	45	72.6	289	2 T20177	hypothetical prote
33	45	72.6	722	2 S57246	ventral nervous sy
34	45	72.6	1414	1 S23809	collagen alpha 2(I)
35	45	72.6	1453	1 S21626	collagen alpha 1(I)
36	45	72.6	1464	1 CGHU1S	collagen alpha 1(I)
37	45	72.6	1669	1 CGHU4B	collagen alpha 3(I)
38	45	72.6	1752	2 A45407	collagen alpha 1(I)
39	45	72.6	1775	2 S31893	collagen alpha 1(I)
40	44	71.0	396	2 T29773	hypothetical prote
41	44	71.0	488	2 A27353	collagen alpha 1(I)
42	44	71.0	673	1 CGB06C	collagen alpha 1(I)
43	44	71.0	680	1 CGHU1D	collagen alpha 1(X)
44	44	71.0	688	2 A53330	collagen alpha 2(I)
45	44	71.0	1707	2 A33526	collagen alpha 2(I)
46	43	69.4	152	2 T24064	hypothetical prote
47	43	69.4	201	2 T46404	hypothetical prote
48	43	69.4	252	2 A55047	hypothetical prote
49	43	69.4	274	2 T20435	hypothetical prote
50	43	69.4	281	2 T32765	hypothetical prote
51	43	69.4	281	2 C88638	protein F58F6.1(I)
52	43	69.4	290	2 A32249	collagen - sea urc
53	43	69.4	294	2 T21668	collagen - nematod
54	43	69.4	295	2 A44984	collagen alpha 1(I)
55	43	69.4	323	2 A61396	collagen alpha 1(I)
56	43	69.4	358	2 T26281	hypothetical prote
57	43	69.4	423	2 A55797	collagen precursor
58	43	69.4	438	2 S53787	collagen alpha cha
59	43	69.4	458	2 T31631	hypothetical prote
60	43	69.4	482	2 B31795	collagen alpha 1(X)
61	43	69.4	614	2 T33149	hypothetical prote
62	43	69.4	623	2 A45137	collagen alpha 4(I)
63	43	69.4	674	2 S13301	collagen alpha 1(X)
64	43	69.4	675	2 S20819	collagen alpha 3(I)
65	43	69.4	744	2 S15435	collagen alpha 1(V)
66	43	69.4	1142	2 JX0369	collagen alpha 1(X)
67	43	69.4	1146	2 A38587	collagen, cornea-s
68	43	69.4	1486	1 B40333	collagen alpha 1(I)
69	43	69.4	1496	1 CGHU2V	collagen alpha 2(V)
70	43	69.4	1497	2 I49607	procollagen type V
71	43	69.4	1691	1 CGHU6B	collagen alpha 6(I)
72	43	69.4	1747	2 A54121	collagen alpha-4 c
73	43	69.4	1763	2 S16366	collagen alpha 2(I)
74	43	69.4	2944	2 A54849	collagen alpha 1(V)
75	42	67.7	206	2 S18250	collagen alpha 1(I)
76	42	67.7	283	2 T19141	hypothetical prote
77	42	67.7	283	2 T32921	hypothetical prote
78	42	67.7	284	2 A41019	serine-trna ligase
79	42	67.7	291	2 T20942	hypothetical prote
80	42	67.7	298	2 JG1448	collagen col-34 -
81	42	67.7	299	2 T29956	hypothetical prote
82	42	67.7	302	2 T15936	hypothetical prote
83	42	67.7	333	2 T23618	hypothetical prote
84	42	67.7	334	2 JG5980	surfactant protein
85	42	67.7	369	2 S33603	ficollin-A precursor
86	42	67.7	370	2 T22510	hypothetical prote
87	42	67.7	371	1 JN0450	conglutinin precu
88	42	67.7	371	2 I45878	Rhs core protein [
89	42	67.7	444	2 G90704	Rhs core protein [
90	42	67.7	444	2 G85554	N-acetylglucosamin
91	42	67.7	486	2 JG7351	collagen alpha 1(I)
92	42	67.7	671	1 CGRT1S	collagen alpha 1(I)
93	42	67.7	779	1 CGB01S	collagen alpha 1(I)
94	42	67.7	931	2 S13580	hypothetical prote
95	42	67.7	963	2 T19140	collagen alpha 2(I)
96	42	67.7	964	1 CGCH2S	collagen alpha 1(V)
97	42	67.7	1019	1 A32856	collagen alpha 1(I)
98	42	67.7	1042	1 CGCH1S	collagen alpha 3(I)
99	42	67.7	1670	1 CGHU3B	collagen alpha 4(I)
100	42	67.7	1690	1 CGHU1V	collagen alpha 4(I)
101	42	67.7	1838	1 CGHU1V	collagen alpha 1(V)
102	42	67.7	1843	2 S18803	collagen alpha 1(V)

103	42	67.7	2551	2	B98047	hypothetical prote	176	39	62.9	290	2	B8638	protein F58F6.2 [1
104	41	66.1	210	2	T20973	hypothetical prote	177	39	62.9	291	2	T20083	hypothetical prote
105	41	66.1	219	2	T14782	hypothetical prote	178	39	62.9	291	2	T26576	hypothetical prote
106	41	66.1	276	2	T19845	hypothetical prote	179	39	62.9	294	2	T29838	hypothetical prote
107	41	66.1	291	2	T34494	hypothetical prote	180	39	62.9	294	2	T29839	hypothetical prote
108	41	66.1	296	2	T24827	hypothetical prote	181	39	62.9	294	2	T22639	hypothetical prote
109	41	66.1	301	2	T23441	hypothetical prote	182	39	62.9	297	2	T18638	hypothetical prote
110	41	66.1	304	2	T16107	hypothetical prote	183	39	62.9	299	2	T22705	hypothetical prote
111	41	66.1	326	2	T16841	hypothetical prote	184	39	62.9	299	2	T24833	hypothetical prote
112	41	66.1	363	2	T16831	hypothetical prote	185	39	62.9	300	2	T24482	hypothetical prote
113	41	66.1	418	2	T15142	hypothetical prote	186	39	62.9	301	2	A53570	collectin-43 - bov
114	41	66.1	486	2	A45137	DNA-binding protei	187	39	62.9	302	2	T21257	hypothetical prote
115	41	66.1	518	2	A55840	macrophage bacteri	188	39	62.9	303	2	T19289	hypothetical prote
116	41	66.1	587	2	AD3519	2,3-dihydroxybenzo	189	39	62.9	304	2	T26185	hypothetical prote
117	41	66.1	645	2	D90782	probable tail fibe	190	39	62.9	305	2	S44767	hypothetical prote
118	41	66.1	645	2	H85642	collagen alpha 1(V	191	39	62.9	310	2	T22641	hypothetical prote
119	41	66.1	744	1	A34246	collagen alpha 1(V	192	39	62.9	310	2	T22641	hypothetical prote
120	41	66.1	744	1	S23298	collagen alpha 1(V	193	39	62.9	313	2	T22828	hypothetical prote
121	41	66.1	775	2	A61228	collagen alpha 2(I	194	39	62.9	329	2	T32783	hypothetical prote
122	41	66.1	917	2	S09646	collagen alpha 2(V	195	39	62.9	348	2	T29288	hypothetical prote
123	41	66.1	1018	1	CGH02A	collagen alpha 2(V	196	39	62.9	375	1	A45225	pulmonary surfacta
124	41	66.1	1029	1	S21369	collagen alpha 2(V	197	39	62.9	380	2	A48295	collagen 1 - marin
125	41	66.1	1532	2	A61262	collagen alpha 1(X	198	39	62.9	381	2	T27806	hypothetical prote
126	41	66.1	1747	2	A45974	collagen alpha 1(X	199	39	62.9	428	2	T24769	hypothetical glyci
127	41	66.1	1758	2	T29350	hypothetical prote	200	39	62.9	439	2	D70954	hypothetical prote
128	41	66.1	1759	2	T29351	collagen alpha 2(I	201	39	62.9	452	2	T30082	hypothetical prote
129	41	66.1	1857	2	S31212	collagen alpha 1(X	202	39	62.9	460	2	T33110	collagen alpha 1(V
130	41	66.1	1888	2	S78476	collagen alpha 1(X	203	39	62.9	576	2	A40695	collagen alpha 1(X
131	41	66.1	3198	2	A43426	collagen alpha 2 f	204	39	62.9	674	2	S23297	hypothetical prote
132	40	64.5	73	2	E70833	hypothetical prote	205	39	62.9	902	2	T26775	collagen alpha 1(I
133	40	64.5	118	2	B41207	collagen 6, nonfib	206	39	62.9	921	2	S42617	collagen alpha 1(V
134	40	64.5	228	2	A44982	collagen UC011 - p	207	39	62.9	1025	2	S34839	collagen alpha 1(V
135	40	64.5	283	2	T29980	hypothetical prote	208	39	62.9	1028	1	CGH01A	collagen alpha 2(I
136	40	64.5	285	2	T29982	hypothetical prote	209	39	62.9	1366	1	CGH02S	collagen alpha 1(X
137	40	64.5	304	2	T22482	hypothetical prote	210	39	62.9	1388	2	A53317	collagen alpha 1(X
138	40	64.5	305	2	T30165	hypothetical prote	211	38.5	62.1	280	2	T03559	2-hydroxyhepta-2,4
139	40	64.5	325	2	S02170	collagen alpha 1(I	212	38	61.3	25	2	I56978	collagen alpha 2(X
140	40	64.5	326	2	B47172	ficolin-beta - pig	213	38	61.3	93	2	I45876	collagen alpha 1(I
141	40	64.5	327	2	T34203	hypothetical prote	214	38	61.3	122	2	A05152	collagen alpha 1(I
142	40	64.5	356	2	S16907	collagen alpha 1(I	215	38	61.3	168	2	PS0036	collagen alpha 1(I
143	40	64.5	357	2	S35250	flagellar biosynth	216	38	61.3	171	2	A34493	collagen alpha 1(I
144	40	64.5	382	2	S20375	collagen alpha 3(V	217	38	61.3	181	2	T13518	collagen alpha 1(I
145	40	64.5	402	1	CGB02S	collagen alpha 2(I	218	38	61.3	210	2	A37358	collagen alpha 1(I
146	40	64.5	464	2	S59513	collagen II A1 pro	219	38	61.3	237	2	A86640	collagen col-7 C
147	40	64.5	469	2	A24450	collagen alpha 2(V	220	38	61.3	245	1	CIHUQA	collagen alpha 1(I
148	40	64.5	473	2	I50629	collagen - chicken	221	38	61.3	248	1	LNH0P1	complement subcomp
149	40	64.5	547	2	A36046	collagen alpha cha	222	38	61.3	248	1	LNH0P1	pulmonary surfacta
150	40	64.5	585	2	S06958	sphingomyelin phos	223	38	61.3	253	2	S49158	complement protei
151	40	64.5	615	2	A05269	collagen alpha 1(I	224	38	61.3	262	2	T42709	hypothetical prote
152	40	64.5	629	1	A39825	sphingomyelin phos	225	38	61.3	275	2	S51641	MYOD protein - rai
153	40	64.5	635	2	A57131	collagen alpha 2(V	226	38	61.3	279	2	T26125	hypothetical prote
154	40	64.5	677	2	S23296	collagen alpha 2(I	227	38	61.3	286	2	S34665	collagen, cuticula
155	40	64.5	707	1	S69781	outer membrane pro	228	38	61.3	287	2	T22637	hypothetical prote
156	40	64.5	742	2	JC7595	scavenger receptor	229	38	61.3	287	2	T15779	hypothetical prote
157	40	64.5	743	1	S23779	collagen alpha 1(V	230	38	61.3	297	2	T18637	hypothetical prote
158	40	64.5	749	2	A70812	hypothetical glyci	231	38	61.3	297	2	T18637	hypothetical prote
159	40	64.5	918	2	S23377	collagen alpha 2(V	232	38	61.3	300	2	T19529	hypothetical prote
160	40	64.5	921	2	S40495	collagen alpha 1(I	233	38	61.3	302	2	T32872	hypothetical prote
161	40	64.5	959	2	S32605	collagen alpha 3(V	234	38	61.3	302	2	A31921	collagen dpy-13 pr
162	40	64.5	1022	2	S04111	collagen alpha 2(V	235	38	61.3	305	2	T20906	hypothetical prote
163	40	64.5	1315	2	A56101	collagen alpha 1(X	236	38	61.3	309	2	T28708	hypothetical prote
164	40	64.5	1373	1	A43291	collagen alpha 2(I	237	38	61.3	310	2	T29731	hypothetical prote
165	40	64.5	1761	2	T13990	collagen type IV a	238	38	61.3	313	2	T33010	hypothetical prote
166	40	64.5	1774	2	B56101	collagen alpha 1(X	239	38	61.3	314	2	T32985	hypothetical prote
167	40	64.5	2274	2	T30258	adenomatous polyop	240	38	61.3	314	2	T32247	hypothetical prote
168	39	62.9	92	2	A38947	oncofetal-laminin	241	38	61.3	316	2	T19291	hypothetical prote
169	39	62.9	107	2	B61396	collagen alpha 1(I	242	38	61.3	316	2	T19288	hypothetical prote
170	39	62.9	187	2	A35980	collagen alpha 1(I	243	38	61.3	316	2	S08169	collagen col-12 pr
171	39	62.9	208	2	T15245	hypothetical prote	244	38	61.3	316	2	S08170	collagen col-13 pr
172	39	62.9	222	2	A88102	protein W09G10.1 [245	38	61.3	318	2	T27977	cuticle collagen d
173	39	62.9	245	1	CIHUQC	complement subcomp	246	38	61.3	337	2	T23794	hypothetical prote
174	39	62.9	266	2	T22706	hypothetical prote	247	38	61.3	348	2	A34705	collagen - Caenorh
175	39	62.9	283	2	T29837	hypothetical prote	248	38	61.3	349	2	S08278	macrophage scaveng

249	38	61.3	356	2	T22827	hypothetical prote	322	36	58.1	253	2	T08152	axonemal inner dyn
250	38	61.3	365	2	S10847	collagen alpha 2(I	323	36	58.1	258	2	B89773	acetoin(diacetyl) r
251	38	61.3	453	2	S08276	macrophage scaveng	324	36	58.1	261	2	A34476	collagen alpha 2(I
252	38	61.3	581	2	I49669	FXR65 binding pro	325	36	58.1	269	2	C83516	hypothetical prote
253	38	61.3	615	2	H70589	hypothetical glyci	326	36	58.1	277	2	JC7903	collectin liver 1
254	38	61.3	632	2	S42731	collagen alpha 1 c	327	36	58.1	285	2	S60598	collagen alpha cha
255	38	61.3	684	2	A30119	collagen alpha 1(X	328	36	58.1	289	2	T26812	hypothetical prote
256	38	61.3	730	2	A36226	collagen alpha 1 c	329	36	58.1	295	2	T22833	hypothetical prote
257	38	61.3	825	1	EDBEXD	immediate-early pr	330	36	58.1	296	2	T21070	hypothetical prote
258	38	61.3	920	2	B34493	collagen alpha 1(I	331	36	58.1	299	2	T20605	hypothetical prote
259	38	61.3	1433	2	A46053	bullous pemphigoid	332	36	58.1	301	2	B31219	collagen 2 - Caeno
260	38	61.3	1545	2	F98262	hypothetical prote	333	36	58.1	304	2	T23801	hypothetical prote
261	38	61.3	1581	2	A30322	glutamate synthase	334	36	58.1	306	2	T21939	hypothetical prote
262	38	61.3	1840	2	T30250	GRI protein - mous	335	36	58.1	311	2	T15268	US10 protein - num
263	37	59.7	138	2	F72732	hypothetical prote	336	36	58.1	312	1	Q0BE07	hypothetical prote
264	37	59.7	140	2	A05249	collagen alpha 1(I	337	36	58.1	312	2	T25048	hypothetical prote
265	37	59.7	177	2	S37749	collagen alpha 2(X	338	36	58.1	316	2	T20497	hypothetical prote
266	37	59.7	178	2	A39762	collagen alpha 1(X	339	36	58.1	317	2	T19143	hypothetical prote
267	37	59.7	244	2	E75635	conserved hypothet	340	36	58.1	323	2	T19142	transforming growt
268	37	59.7	246	2	S29328	complement subcomp	341	36	58.1	323	2	A47172	hypothetical prote
269	37	59.7	247	1	LNRPBS	pulmonary surfacta	342	36	58.1	324	2	T18763	collagen sqt-1 pre
270	37	59.7	248	1	LNHUPS	pulmonary surfacta	343	36	58.1	324	2	A31920	extracellular este
271	37	59.7	248	1	LNHUP6	pulmonary surfacta	344	36	58.1	326	2	JC6022	hypothetical prote
272	37	59.7	248	1	LNRTPS	pulmonary surfacta	345	36	58.1	330	2	T26004	hypothetical prote
273	37	59.7	248	2	I51921	pulmonary surfacta	346	36	58.1	337	2	T21055	gene V protein - p
274	37	59.7	278	2	S44796	F09G8.6 protein -	347	36	58.1	340	1	B46345	hypothetical prote
275	37	59.7	286	2	B45632	merozoite surface	348	36	58.1	345	2	T29981	surfactant protein
276	37	59.7	294	2	H75080	probable homoserin	349	36	58.1	374	1	A42046	dnaj protein homol
277	37	59.7	299	2	T19564	hypothetical prote	350	36	58.1	381	2	T43517	hypothetical prote
278	37	59.7	299	2	T25407	hypothetical prote	351	36	58.1	386	2	JC7356	hypothetical prote
279	37	59.7	307	2	T18846	hypothetical prote	352	36	58.1	390	2	JC7256	L-pipecolate oxida
280	37	59.7	307	2	T19582	hypothetical prote	353	36	58.1	402	1	QB072	microtubule-associ
281	37	59.7	314	2	T34269	hypothetical prote	354	36	58.1	428	2	S35215	sodium channel pro
282	37	59.7	319	2	T62227	hypothetical prote	355	36	58.1	438	2	D90734	probable tail fibe
283	37	59.7	321	2	T28760	hypothetical prote	356	36	58.1	448	1	QB071	microtubule-associ
284	37	59.7	324	2	T28032	hypothetical prote	357	36	58.1	461	2	F70571	collagen alpha 3(I
285	37	59.7	326	2	S61517	ficolin-1 precursor	358	36	58.1	471	2	A39024	hypothetical prote
286	37	59.7	327	2	T29031	hypothetical prote	359	36	58.1	482	2	S76475	gas vesicle protei
287	37	59.7	330	2	S46657	collagen alpha 1(X	360	36	58.1	492	2	S28139	hypothetical prote
288	37	59.7	333	2	T20436	hypothetical prote	361	36	58.1	518	2	T23745	hypothetical prote
289	37	59.7	341	2	T16296	hypothetical prote	362	36	58.1	534	2	F90031	PTS system, arbuti
290	37	59.7	360	2	T37285	collagen dpy-2 - C	363	36	58.1	536	2	S15183	gas-vesicle operon
291	37	59.7	366	2	S11449	collagen short cha	364	36	58.1	536	2	T08241	gas-vesicle operon
292	37	59.7	367	2	JC4831	adsorption protein	365	36	58.1	545	2	S28117	gas-vesicle operon
293	37	59.7	380	2	T28888	cuticle collagen d	366	36	58.1	571	2	S52750	alpha-amylase fami
294	37	59.7	394	2	T33641	hypothetical prote	367	36	58.1	580	2	B87532	dehri protein - Pa
295	37	59.7	404	2	C75027	dihydroorotase (py	368	36	58.1	588	2	F70971	hypothetical glyci
296	37	59.7	423	2	A41207	collagen 13, nonfi	369	36	58.1	588	2	T25835	hypothetical prote
297	37	59.7	427	2	T20800	hypothetical prote	370	36	58.1	599	2	T08613	hypothetical prote
298	37	59.7	435	2	T15143	hypothetical prote	371	36	58.1	712	2	B87253	polyribonucleotide
299	37	59.7	453	2	S18804	collagen alpha 4(I	372	36	58.1	712	2	B83067	polyribonucleotide
300	37	59.7	471	2	S15035	acetylcholinestera	373	36	58.1	713	2	A25586	polyribonucleotide
301	37	59.7	487	2	JC7126	testis zinc finger	374	36	58.1	714	2	AC3497	inorganic diphosph
302	37	59.7	518	2	T49778	hypothetical prote	375	36	58.1	771	2	S72526	conserved hypothet
303	37	59.7	624	2	A55576	collagen alpha 2(X	376	36	58.1	827	2	T40394	hypothetical glyci
304	37	59.7	783	2	F70824	hypothetical glyci	377	36	58.1	837	2	E70835	protein-tyrosine k
305	37	59.7	914	2	H70987	hypothetical glyci	378	36	58.1	876	2	A49508	receptor tyrosine
306	37	59.7	992	2	T08772	hypothetical prote	379	36	58.1	913	2	A48280	hypothetical colle
307	37	59.7	1051	2	A35763	collagen alpha 2 c	380	36	58.1	929	2	T38948	hypothetical glyci
308	37	59.7	1120	2	H88449	protein F54D8.1 (i	381	36	58.1	929	2	E70917	sodium channel pro
309	37	59.7	1616	2	E90704	Rhs core protein w	382	36	58.1	1329	2	T37098	gelatin factor AB
310	37	59.7	1645	2	H85554	hypothetical prote	383	36	58.1	2108	2	A37098	collagen alpha 3(V
311	37	59.7	3124	2	A40020	collagen alpha 1(X	384	36	58.1	2647	2	A37797	collagen alpha 3(V
312	37	59.7	3436	2	S55659	tegument protein 6	385	36	58.1	3137	2	CGH3A	collagen alpha 3(V
313	36	58.1	35	2	B24450	collagen alpha 1(V	386	36	58.1	3176	2	CGH3A	collagen alpha 3(V
314	36	58.1	48	2	B05249	collagen alpha 2(I	387	35	56.5	53	2	I60384	gene 11 protein -
315	36	58.1	99	2	I57012	alpha 2(XI) collag	388	35	56.5	58	2	AF1845	hypothetical prote
316	36	58.1	141	2	B98289	hypothetical prote	389	35	56.5	74	2	T16979	metallothionein-11
317	36	58.1	141	2	AH2994	hypothetical prote	390	35	56.5	86	2	S00802	collagen-like 1(X
318	36	58.1	184	1	CGK72S	collagen alpha 2(I	391	35	56.5	86	1	CGBEHS	collagen-like prot
319	36	58.1	212	2	JC7511	fibroblast growth	392	35	56.5	99	2	T32603	hypothetical prote
320	36	58.1	253	1	ClHUOB	complement subcomp	393	35	56.5	102	2	T31879	hypothetical prote
321	36	58.1	253	2	I49560	complement Clq B C	394	35	56.5	108	2	A46222	hydrophobin ceg-2

F:197-358/Domain: collagenous COL2 #status predicted <COL2>
F:359-380/Domain: non-collagenous NC3 #status predicted <NC3>
F:381-615/Domain: collagenous COL3 #status predicted <COL3>
F:616-633/Domain: non-collagenous NC4 #status predicted <NC4>

Query Match 93.5% Score 58; DB 2; Length 633;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGL 11
|||||
Db 270 GEKGAEGSPGL 280

RESULT 2
S23810
collagen alpha 1(XVI) chain precursor - human
N:Alternate names: procollagen alpha 1(XVI) chain
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence revision 28-Oct-1994 #text_change 09-Jul-2004
C:Accession: S23810; PQ0612; S08012
R:Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.
Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992
A:Title: Cloning and chromosomal location of human alpha1(XVI) collagen.
A:Reference number: S23810; MUID:92335339; PMID:1631157
A:Accession: S23810
A:Molecule type: mRNA
A:Residues: 1-1603 <PAN>
A:Cross-references: UNIPROT:Q07092; UNIPARC:UPI0000126D33; EMBL:M92642; NID:G180757; PID
A:Experimental source: skin fibroblasts
R:Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi
J. Biochem. 112, 856-863, 1992
A:Title: Molecular cloning and partial characterization of a novel collagen chain, alpha
A:Reference number: PQ0612; MUID:93203161; PMID:1284248
A:Accession: PQ0612
A:Molecule type: mRNA
A:Residues: 'GGR', 421-536, 'P', 538-1159, 'P', 1161-1162, 'P', 1164, 'P', 1166-1603 <YAM>
A:Cross-references: UNIPARC:UPI000016B3C2; GB:S57132; NID:G298641; PIDN:AAB25797.1; PID
A:Experimental source: placenta
R:Kimura, S.
submitted to the EMBL Data Library, April 1989
A:Description: Partial nucleotide and amino acid sequence of a collagen-like protein fr
A:Reference number: S08012
A:Accession: S08012
A:Molecule type: mRNA
A:Residues: 403-419, 'GR', 421-536, 'P', 538-846, 'VM' <KIM>
A:Cross-references: UNIPARC:UPI000073DAB; EMBL:X14963; NID:G299984; PIDN:CAA33085.1; PID
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL16A1
A:Cross-references: GDB:134045; OMIM:120326
A:Map position: lp34-lp34
A:Complex: type XVI collagen may be a homotrimer, or a heterotrimer of two alpha 1(XVI)
C:Function:
A:Description: structural component of extracellular fibrous polymer as a minor form pro
A:Note: may play a role in forming elastic connections at fibril surfaces
C:Keywords: Cell binding; coiled coil; extracellular matrix; glycoprotein; hydroxyllysine
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-1603/Product: collagen alpha 1(XVI) chain #status predicted <MAT>
F:22-333/Domain: amino-terminal nonhelical #status predicted <NC11>
F:334-1577/Region: interrupted helical
F:334-360/Domain: collagenous COL10 #status predicted <COL0>
F:375-505/Domain: collagenous COL9 #status predicted <COL9>
F:521-554/Domain: collagenous COL8 #status predicted <COL8>
F:539-541/Region: cell attachment (R-G-D) motif
F:572-630/Domain: collagenous COL7 #status predicted <COL7>
F:652-722/Domain: collagenous COL6 #status predicted <COL6>
F:738-875/Domain: collagenous COL5 #status predicted <COL5>
F:887-938/Domain: collagenous COL4 #status predicted <COL4>
F:973-987/Domain: collagenous COL3 #status predicted <COL3>
F:1005-1007/Region: cell attachment (R-G-D) motif
F:1011-1432/Domain: collagenous COL2 #status predicted <COL2>

F;1226-1228/Region: cell attachment (R-G-D) motif
 F;1472-1577/Domain: collagenous COL1 #status predicted <COL1>
 F;1578-1603/Domain: carboxyl-terminal nonhelical #status predicted <NCOL1>
 F;47,327/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 83.9%; Score 52; DB 2; Length 1603;
 Best Local Similarity 81.8%; Pred. No. 1.1;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPL 11
 |||||:||||:
 Db 817 GKGAGSGPGV 827

RESULT 3

S32436
 collagen alpha 2(IX) chain - human (fragment)
 C;Species: Homo sapiens (man)
 C;Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 15-Sep-2003
 C;Accession: S32436; S34487; S64673
 R;Peraelae, M.; Haenninen, M.; Haestbacka, J.; Elima, K.; Vuorio, E.
 FEBS Lett. 319, 177-180, 1993
 A;Title: Molecular cloning of the human alpha-2(IX) collagen cDNA and assignment of the
 A;Reference number: S32436; MUID:93202262; PMID:8454052
 A;Accession: S32436

A;Molecule type: mRNA
 A;Residues: 1-618 <PER1>
 A;Cross-references: UNIPARC:UPI000017A166; EMBL:M95610; NID:G1054872
 R;Peraelae, M.; Haenninen, M.; Haestbacka, J.; Vuorio, E.
 submitted to the EMBL Data Library, March 1993
 A;Description: Molecular cloning of the human alpha-2 (IX) collagen cDNA and assignment
 A;Reference number: S34487
 A;Accession: S34487
 A;Molecule type: mRNA
 A;Residues: 1-26,'QT','29','S','31-32','LM','35-561','L','563-578','P','580-618 <PER2>
 A;Cross-references: UNIPARC:UPI000017A167; EMBL:M95610; NID:G1054872
 R;Diab, M.; Wu, J.J.; Eyre, D.R.
 Biochem. J. 314, 327-332, 1996

A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
 A;Reference number: S64673; MUID:96195147; PMID:8660302
 A;Accession: S64673
 A;Molecule type: protein
 A;Residues: 123-133,'P',135-137 <DFA>
 A;Cross-references: UNIPARC:UPI000017A168
 C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C;Genetics:

A;Gene: GDB:COL9A2
 A;Cross-references: GDB:138310; OMIM:120260
 A;Map position: lp33-lp32.2
 C;Complex: type IX collagen may be a heterotrimer of one alpha 1(IX) chain, one alpha 2
 C;Function:
 A;Description: structural component of extracellular fibrous polymer associated with typ
 A;Keywords: chondroitin sulfate proteoglycan; coiled coil; extracellular matrix; glycop
 F;1-114/Domain: collagenous COL3 (fragment) #status predicted <COL3>
 F;115-131/Domain: non-collagenous NC3 #status predicted <NC3>
 F;132-470/Domain: collagenous COL2 #status predicted <COL2>
 F;471-500/Domain: non-collagenous NC2 #status predicted <NC2>
 F;501-615/Domain: collagenous COL1 #status predicted <COL1>
 F;616-618/Domain: non-collagenous NC1 (fragment) #status predicted <NC1>
 F;120/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 80.6%; Score 50; DB 2; Length 618;
 Best Local Similarity 81.8%; Pred. No. 0.96;
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPL 11
 |||||:||||:
 Db 228 GKGDESGPI 238

RESULT 4

T22002

hypothetical protein F39H11.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T22002
 R;White, S.

submitted to the EMBL Data Library, October 1996
 A;Reference number: Z19500
 A;Accession: T22002
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-650 <WIL>
 A;Cross-references: UNIPROT:O17866; UNIPARC:UPI000006101E; EMBL:Z81079; PIDN:CAB03084.1;
 A;Experimental source: clone F39H11

C;Genetics:

A;Gene: CRSP:F39H11.4
 A;Map position: 1
 A;Introns: 109/3; 154/1; 357/1; 420/3; 464/3; 566/2; 594/1; 628/3

Query Match 79.0%; Score 49; DB 2; Length 650;
 Best Local Similarity 72.7%; Pred. No. 1.5;
 Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPL 11
 |||||:||||:
 Db 206 GKGADGAPCL 216

RESULT 5

S31216
 collagen alpha 1(X) chain precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 30-Sep-1993 #sequence revision 30-Sep-1993 #text_change 09-Jul-2004
 C;Accession: S31216; S28807; S22215; S30127; I48299; S26397; S31830
 R;Kong, R.Y.C.; Kwan, K.M.; Lau, E.T.; Thomas, J.T.; Boot-Handford, R.P.; Grant, M.E.; C
 Eur. J. Biochem. 213, 99-111, 1993
 A;Title: Intron-exon structure, alternative use of promoter and expression of the mouse
 A;Reference number: S31216; MUID:93238750; PMID:8477738
 A;Accession: S31216

A;Molecule type: DNA
 A;Residues: 1-680 <KON>
 A;Cross-references: UNIPROT:Q05306; UNIPARC:UPI000016CBAA; EMBL:Z21610; NID:949793; PIDN
 R;Elima, K.; Metsaeranta, M.; Kallio, J.; Peraelae, M.; Eroola, I.; Garofalo, S.; de Cro
 Biochem. J. 289, 247-253, 1993
 A;Title: The mouse collagen X gene: complete nucleotide sequence, exon structure and exp
 A;Reference number: S28807; MUID:93143676; PMID:8424763
 A;Accession: S28807

A;Molecule type: DNA
 A;Residues: 1-285,'A','287-680 <ELI>

A;Cross-references: UNIPARC:UPI000026913; EMBL:X67348; NID:950480; PIDN:CAA47763.1; PID
 R;Elima, K.; Metsaeranta, M.; Kallio, J.; Peraelae, M.; Eroola, I.; Garofalo, S.; de Cro
 Biochim. Biophys. Acta 1130, 78-80, 1992

A;Title: Specific hybridization probes for mouse alpha-2(IX) and alpha-1(X) collagen mRN
 A;Reference number: S22215; MUID:92182017; PMID:1543751

A;Accession: S22215

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 385-450,'K',452-627 <ELA>

A;Cross-references: UNIPARC:UPI000016CBAB; EMBL:X63013; NID:949795; PIDN:CAA44741.1; PID
 R;Apte, S.S.; Olsen, B.R.

Matrix 13, 165-179, 1993

A;Title: Characterization of the mouse type X collagen gene.

A;Reference number: S30127; MUID:93261348; PMID:8492743

A;Accession: S30127

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-12,'F',14-26,'S',28-247,'L',249-285,'A',287-305,'F',307-416,'S',418-499,'L'
 A;Cross-references: UNIPARC:UPI000017385

R;Apte, S.S.; Seidin, M.F.; Hayashi, M.; Olsen, B.R.

Eur. J. Biochem. 206, 217-224, 1992

A;Title: Cloning of the human and mouse type X collagen genes and mapping of the mouse t
 A;Reference number: I48299; MUID:92267014; PMID:1587271

A;Accession: I48299

A;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 52-247, 'L', 249-285, 'A', 287-305, 'P', 307-416, 'S', 418-499, 'L', 501-566, 'C', 568, 'A';
A:Cross-references: UNIPARC:UPI000016CCAC; EMBL:X65121; NID:G50482; PIDN:CAA46237.1; PID
R:Summers, T.A.; Irwin, M.H.; Mayne, R.; Ballian, G.
J. Biol. Chem. 263, 581-587, 1988
A:Title: Monoclonal antibodies to type X collagen. Biosynthetic studies using an antibody
A:Reference number: S26397; MUID:88087150; PMID:2826450
A:Accession: S26397
A:Molecule type: protein
A:Residues: 'SPGYFSQ', 24-26, 'KQ', <SUM>
A:Cross-references: UNIPARC:UPI0000173B6
C:Genetics:
A:Gene: Col10a-1
A:Map position: 10
A:Introns: 51/3
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology
C:Keywords: coiled coll; extracellular matrix; glycoprotein; homotrimer
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-680/Product: collagen alpha 1(X) chain #status predicted <MAT>
F:553-679/Domain: complement C1q carboxyl-terminal homology <CIQ>
Query Match 79.0%; Score 49; DB 2; Length 680;
Best Local Similarity 72.7%; Pred. No. 1.6;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 GEKGAGSGPGL 11
|||||:|:
DB 174 GEKGAGGAPGV 184
RESULT 6
CGH2B
collagen alpha 2(IV) chain precursor - human
N:Alternate names: procollagen alpha 2(IV) chain
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence revision 03-Oct-1995 #text change 09-Jul-2004
C:Accession: A32024; S00007; S02624; S00246; S17678; S16911; B32117; S16877; S00165; S39
R:Hostikka, S.L.; Tryggvason, K.
J. Biol. Chem. 263, 19488-19493, 1988
A:Title: The complete primary structure of the alpha2 chain of human type IV collagen an
A:Reference number: A32024; MUID:89066769; PMID:3198637
A:Accession: A32024
A:Molecule type: mRNA
A:Residues: 1-1712 <HOS1>
A:Cross-references: UNIPROT:P08572; UNIPARC:UPI0000126D42; EMBL:J04210; EMBL:X05610; GB:
R:Hostikka, S.L.; Kurkinen, M.; Tryggvason, K.
FEBS Lett. 216, 281-286, 1987
A:Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA re
ated region.
A:Reference number: S00007; MUID:87219158; PMID:3582677
A:Accession: S00007
A:Molecule type: mRNA
A:Residues: 1254-1398, 'V', 1400-1712 <HOS2>
A:Cross-references: UNIPARC:UPI0000173BE6; EMBL:J04210; EMBL:X05610; GB:M20753; NID:G295
A:Note: 1399-1le was also found
R:Hostikka, S.L.; Tryggvason, K.
FEBS Lett. 224, 297-305, 1987
A:Title: Extensive structural differences between genes for the alpha(1) and alpha(2) ch
A:Reference number: S02624; MUID:88083553; PMID:2826228
A:Accession: S02624
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1347-1350, 1377-1393, 1426-1432, 1465-1471, 1529-1535, 1625-1630 <HOS3>
A:Cross-references: UNIPARC:UPI0000142E66; UNIPARC:UPI0000173BE7; UNIPARC:UPI0000173BE8;
A:Note: complete nucleotide sequence not shown
R:Brazel, D.; Pollner, R.; Oberbauer, I.; Kuehn, K.
Eur. J. Biochem. 172, 35-42, 1988
A:Title: Human basement membrane collagen (type IV): the amino acid sequence of the alph
A:Reference number: S00246; MUID:88151998; PMID:3345760
A:Accession: S00246
A:Molecule type: mRNA
A:Residues: 1-682, 'G', 684-1043 <BRA>
A:Cross-references: UNIPARC:UPI0000173BEC; EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID

R:Oberbauer, I.
submitted to the EMBL Data Library, June 1987
A:Reference number: S17678
A:Accession: S17678
A:Molecule type: mRNA
A:Residues: 1-470, 'P', 472-682, 'G', 684-1043 <OBE>
A:Cross-references: UNIPARC:UPI000016A709; EMBL:X05562; NID:G30075; PIDN:CAA29076.1; PID
R:Poeschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A:Title: The genes for the alpha(IV) and alpha2(IV) chains of human basement membrane c
A:Reference number: S02738; MUID:89030632; PMID:2846280
A:Accession: S16911
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-33 <POE>
A:Cross-references: UNIPARC:UPI000016A6F3; EMBL:X12784; GB:M36963; NID:G30072; PIDN:CAA3
R:Soiininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A:Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A:Reference number: A92690; MUID:89034231; PMID:3182844
A:Accession: B32117
A:Molecule type: DNA
A:Residues: 1-33 <SO11>
A:Cross-references: UNIPARC:UPI000016A6F3; EMBL:J04217; EMBL:J05039; NID:G180759; PIDN:A
R:Soiininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A:Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
A:Reference number: S16876; MUID:89340433; PMID:2701944
A:Accession: S16877
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-33 <SO12>
A:Cross-references: UNIPARC:UPI000016A6F3; EMBL:J04217; NID:G180759; PIDN:AAAS3097.1; PI
A:Note: this sequence was submitted to the EMBL Data Library, October 1988
R:Siebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, H.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 168, 569-575, 1987
A:Title: Construction of a model for the aggregation and cross-linking region (7S domain
is region.
A:Reference number: S00165; MUID:88029476; PMID:3117548
A:Accession: S00165
A:Molecule type: protein
A:Residues: 37-247 <SIE1>
A:Cross-references: UNIPARC:UPI0000173BED
A:Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-Gly
R:Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen
A:Reference number: S39614; MUID:94038963; PMID:8223488
A:Accession: S39615
A:Molecule type: protein
A:Residues: 407-570 <EBL>
A:Cross-references: UNIPARC:UPI0000173BEE
R:Macwright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (r
A:Reference number: S16910; MUID:84053346; PMID:6416291
A:Accession: S16912
A:Molecule type: protein
A:Residues: 490-492, 'X', 494-496, 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>
A:Cross-references: UNIPARC:UPI0000173BEF; UNIPARC:UPI0000173BF0
A:Experimental source: Placenta
R:Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A:Title: Pepsin fragments of human placental basement-membrane collagens showing inter
A:Reference number: S16908; MUID:82005835; PMID:6792033
A:Accession: B58517
A:Molecule type: protein
A:Residues: 490-492, 'X', 494-501, 'P', 503-507, 952-957, 'X', 959-966, 'X', 968, 984-986, 'X', 988
81-1195 <GLA>
A:Cross-references: UNIPARC:UPI0000173BF1; UNIPARC:UPI0000173BF2; UNIPARC:UPI0000173BF3;
R:Killen, P.D.; Francomano, C.A.; Yamada, Y.; Modi, W.S.; O'Brien, S.J.
Hum. Genet. 77, 318-324, 1987
A:Title: Partial structure of the human alpha-2(IV) collagen chain and chromosomal local

A;Reference number: S01450; MUID:88085168; PMID:3692475
A;Accession: S01450
A;Molecule type: mRNA
A;Residues: 1040,'L',1042-1398,'V',1400-1418,'M',1420-1635,'V',1637-1712 '<KIL>
A;Cross-references: UNIPARC:UPI0000072E68; EMBL:M24766; NID:9537328; PIDN:AAA52043.1; PT
R;Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A;Reference number: S02550; MUID:89005112; PMID:2844531
A;Accession: S02550
A;Molecule type: protein
A;Residues: 1480-1535;1545-1614;1617-1662,'H',1664-1700,'G';1705-1708;1710-1712 '<SIB2>
A;Cross-references: UNIPARC:UPI0000173BF6; UNIPARC:UPI0000173BF7; UNIPARC:UPI0000173BF8;
A;Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17
R;Myers, J.C.; Howard, P.S.; Gelen, A.M.; Dion, A.S.; Macarak, E.J.
J. Biol. Chem. 262, 9231-9238, 1987
A;Title: Duplication of type IV collagen COOH-terminal repeats and species-specific exp
A;Reference number: A27114; MUID:87250571; PMID:2439508
A;Accession: B27114
A;Molecule type: mRNA
A;Residues: 1486-1574,'I',1576-1712 '<MYE>
A;Cross-references: UNIPARC:UPI0000173BF6; EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PT
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A2
A;Map position: 13q34-13q34
A;Cross-references: GDB:119792; OMIM:120090
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete
A;Note: the alpha 1(IV) and alpha 2(IV) chain genes are encoded on opposite strands with
C;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CGHU4B)
domains (with disulfide and desmosine cross-links), dimeric associations among trimer ca
rupted helical domain (with disulfide and desmosine cross-links).
C;Function:
A;Description: structural component of basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glycopro
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-1712/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F;29-57/Domain: amino-terminal nonhelical, NH1 <NH1>
F;58-1485/Region: interrupted helical
F;362-364/Region: cell attachment (R-G-D) motif
F;784-786/Region: cell attachment (R-G-D) motif
F;868-870/Region: cell attachment (R-G-D) motif
F;889-891/Region: cell attachment (R-G-D) motif
F;970-972/Region: cell attachment (R-G-D) motif
F;1069-1071/Region: cell attachment (R-G-D) motif
F;1228-1230/Region: cell attachment (R-G-D) motif
F;1452-1454/Region: cell attachment (R-G-D) motif
F;1486-1712/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1495-1593/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1603-1708/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;42,47,51,53,137,483,485/Disulfide bonds: interchain #status predicted
F;57,87,90,102,165,168,225,239,242/Binding site: carboxylate (Lys) (covalent) #status p
F;57/Modified site: 5-hydroxylysine (Lys) #status atypical
F;83,75,96,114,120,123,132,150,159,186,189,201,213,216,219,496,499,955,964,1103,1115
F;87,90,102,165,168,225,239,242/Modified site: 5-hydroxylysine (Lys) #status experimental
F;138/Binding site: carboxylate (Asn) (covalent) #status experimental
F;209/Modified site: 4-hydroxyproline (Pro) #status atypical
F;661-681/Disulfide bonds: #status predicted
F;1275/Binding site: carboxylate (Asn) (covalent) #status predicted
F;1504-1590,1537-1593/Disulfide bonds: (or 1504-1593, 1537-1590) #status experimental
F;1549-1555,1658-1665/Disulfide bonds: #status experimental
F;1612-1705,1646-1708/Disulfide bonds: (or 1612-1708, 1646-1705) #status experimental

Query Match 79.0%; Score 49; DB 1; Length 1712;
Best Local Similarity 72.7%; Pred. No. 3.9;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKAGGSGFL 11
|||||
Db 276 GEKSGEGPGI 286

RESULT 7

S41067

collagen alpha 1(III) chain - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 13-Jan-1995 #sequence revision 13-Jan-1995 #text_change 09-Jul-2004

C;Accession: S41067; A29905; S31924

R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.

Biochim. Biophys. Acta 1217, 41-48, 1994

A;Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa

A;Reference number: S41067; MUID:94114571; PMID:8286415

A;Accession: S41067

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-636 <GLU>

A;Cross-references: UNIPROT:P13941; UNIPARC:UPI0000126D1A; EMBL:X70369; NID:957915; PIDN

R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Komm, B.; Mohn, K.

DNA 7, 347-354, 1988

A;Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2 (V) collagen mRNAs by estr

A;Reference number: A29905; MUID:88296083; PMID:2456904

A;Accession: A29905

A;Molecule type: mRNA

A;Residues: 308-482 <FRA>

A;Cross-references: UNIPARC:UPI0000170BC0; GB:M21354; NID:9203500; PIDN:AAA40942.1; PID:

R;Glumoff, V.; Maekelae, J.K.; Vuorio, E.

submitted to the EMBL Data Library, February 1993

A;Reference number: S31924

A;Accession: S31924

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 2-636 <GL2>

A;Cross-references: UNIPARC:UPI000017737E; EMBL:X70369

C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

F;408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 77.4%; Score 48; DB 2; Length 636;

Best Local Similarity 80.0%; Pred. No. 2.2;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKAGGSGPG 10

|||||

Db 342 GERGSESGPG 351

RESULT 8

CGB07S

collagen alpha 1(III) chain - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 09-Jul-2004

C;Accession: A02862; A38001; A38002; A38003; A38004; A38005; S71946

R;Fietzek, P.P.; Allmann, H.; Rauterberg, J.; Henkel, W.; Wächter, E.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979

A;Title: The covalent structure of calf skin type III collagen. I. The amino acid sequen

A;Reference number: A02862; MUID:80026026; PMID:488906

A;Accession: A02862

A;Molecule type: protein

A;Residues: 1-242 <FIE>

A;Cross-references: UNIPROT:P04258; UNIPARC:UPI0000173B8A

R;Dewes, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979

A;Title: The covalent structure of calf skin type III collagen. II. The amino acid sequ

A;Reference number: A38001; MUID:80026027; PMID:488907

A;Accession: A38001

A;Molecule type: protein

A;Residues: 243-422 <DEW1>

A;Cross-references: UNIPARC:UPI0000173B8B

R;Bentz, H.; Fietzek, P.P.; Kuehn, K.

Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979

A;Title: The covalent structure of calf skin type III collagen. III. The amino acid sequ

A;Reference number: A38002; MUID:80026028; PMID:488908

A;Accession: A38002

A;Molecule type: protein

A;Residues: 423-571 <BEN>
A;Cross-references: UNIPARC:UPI0000173B8C
R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979
A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid sequence
A;Reference number: A38003; MUID:80026029; PMID:488909
A;Accession: A38003
A;Molecule type: protein
A;Residues: 572-808 <LAN>
A;Cross-references: UNIPARC:UPI0000173B8D
R;Dewes, H.; Fietzek, P.P.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979
A;Title: The covalent structure of calf skin type III collagen. V. The amino acid sequence
A;Reference number: A38004; MUID:80026030; PMID:488910
A;Accession: A38004
A;Molecule type: protein
A;Residues: 809-947 <DEW>
A;Cross-references: UNIPARC:UPI0000173B8E
R;Allmann, H.; Fietzek, P.P.; Glanville, R.W.; Kuehn, K.
Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979
A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid sequence
A;Reference number: A38005; MUID:80026031; PMID:488911
A;Accession: A38005
A;Molecule type: protein
A;Residues: 948-1049 <ALL>
A;Cross-references: UNIPARC:UPI0000173B8F
A;Experimental source: skin
R;Henkel, W.
Biochem. J. 318, 497-503, 1996
A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen.
A;Reference number: S71946; MUID:96404897; PMID:8809038
A;Accession: S71946
A;Molecule type: protein
A;Residues: 87-106;1017-1029;1037-1049 <HEN>
A;Cross-references: UNIPARC:UPI0000173B90; UNIPARC:UPI0000173B91; UNIPARC:UPI0000173B92
C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are highly conserved.
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-1049/Product: collagen alpha 1(I) chain #status experimental <CAB>
F;1-14/Region: amino-terminal nonhelical telopeptide
F;15-1040/Region: helical
F;587-589/Region: cell attachment (R-G-D) motif
F;752-754/Region: cell attachment (R-G-D) motif
F;875-877/Region: cell attachment (R-G-D) motif
F;878-880/Region: cell attachment (R-G-D) motif
F;935-937/Region: cell attachment (R-G-D) motif
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide
F;95,107,119,938,950/Modified site: 5-hydroxylysine (Lys) #status experimental
F;107,950/Modified site: allysine (Lys) #status predicted
F;107/Binding site: carboxylate (Lys) (covalent) #status experimental
F;1040,1041/Disulfide bonds: interchain #status predicted

Query Match 77.4%; Score 48; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred.No. 3.6;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGEGSPG 10
Db 1014 GERGEGSPG 1023

RESULT 9
S59856
collagen alpha 1(III) chain precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 09-Jul-2004
C;Accession: S59856; S62120; S16373
R;Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA
A;Reference number: S59856; MUID:95011609; PMID:7926795
A;Accession: S59856

Biochemistry 16, 7201, 1988
A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pro-
A;Reference number: S01726; MUID:98303360; PMID:3405773
A;Accession: S01726
A;Molecule type: mRNA
A;Residues: 1-170 <TM>
A;Cross-references: UNIPARC:UPI000016A706; EMBL:X07240; NID:G30060; PIDN:CAA30229.1; PID:
A;Note: the authors translated the codon CAG for residue 154 as His
R;Janeczko, R.A.; Ramirez, F., 1989
Nucleic Acids Res. 17, 6742, 1989
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
A;Reference number: S04987; MUID:89386015; PMID:2780304
A;Accession: S04987
A;Molecule type: mRNA
A;Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634,
A;Cross-references: UNIPARC:UPI000016A61C; EMBL:X13332; NID:G29545; PIDN:CAA3387.1; PID:
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide
R;Seyer, J.M.; Kang, A.H.,
Biochemistry 16, 1158-1164, 1977
A;Title: Covalent structure of collagen: amino acid sequence of cyanogen bromide peptide
A;Reference number: A90399; MUID:77134724; PMID:557335
A;Accession: A90399
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-232, 'P', 234-292, 'D', 294-398 <SEY1>
A;Cross-references: UNIPARC:UPI0000173B81
A;Experimental source: liver
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-galact
submitted to the Atlas, December 1977
A;Reference number: A94562
A;Accession: A94562
A;Molecule type: protein
A;Residues: 'V', 169-225, 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>
A;Cross-references: UNIPARC:UPI0000173B82
A;Experimental source: liver
A;Note: author submitted corrections to A90399
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.H.
Am. J. Hum. Genet. 53, 62-70, 1993
A;Title: Parental somatic and germ-line mosaicism for a multixon deletion with unusual
ispring.
A;Reference number: I51868; MUID:93304430; PMID:8317500
A;Accession: I51868
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 186-194 <ML>
A;Cross-references: UNIPARC:UPI0000000B14; GB:SG2925; NID:G386425; PIDN:AAD13937.1; PID:
R;Chiodo, A.A.; Sillence, D.O.; Cole, W.G.; Bateman, J.F.
Biochem. J. 311, 939-943, 1995
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the COL3
A;Reference number: S59511; MUID:96067614; PMID:7487954
A;Accession: S59511
A;Molecule type: mRNA
A;Residues: 302-423 <CHI>
A;Cross-references: UNIPARC:UPI0000173B83; GB:S79877; NID:G1195576; PIDN:AAB35615.1; PID:
R;Seyer, J.M.; Kang, A.H., 1978
Biochemistry 17, 3404-3411, 1978
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr pe
A;Reference number: A90414; MUID:79000343; PMID:687591
A;Accession: A90414
A;Molecule type: protein
A;Residues: 399-675, 'N', 677-727 <SEY3>
A;Cross-references: UNIPARC:UPI0000173B84
A;Experimental source: liver
R;Lee, B.; Vitale, E.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.
J. Biol. Chem. 266, 5256-5259, 1991
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of th
A;Reference number: I55349; MUID:91161621; PMID:1672129
A;Accession: I55349
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 537-605 <LEE>
A;Cross-references: UNIPARC:UPI0000000A2; GB:M59312; NID:G180815; PIDN:AAA52041.1; PID:
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.

Biochemistry 19, 1583-1589, 1980
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from tvi
A;Reference number: A90438; MUID:80198282; PMID:6246925
A;Accession: A90438
A;Molecule type: protein
A;Residues: 728-895, 'A', 897-964 <SEY4>
A;Cross-references: UNIPARC:UPI0000173B85
A;Experimental source: liver
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; Chan,
J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and
A;Reference number: A38303; MUID:91009133; PMID:2145268
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: UNIPARC:UPI0000000A41; GB:J05617; GB:M55603; GB:M59227; NID:G180878;
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos synd
R;Mankoo, B.S.; Dalgleish, R.,
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: UNIPARC:UPI0000173B86; EMBL:X06700; NID:G30053; PIDN:CAA29886.1; PID:
R;Seyer, J.M.; Kang, A.H.,
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from tvi
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1
A;Cross-references: UNIPARC:UPI0000173B87
A;Experimental source: liver
R;Loiold, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenbloom, J.; Myer
Nucleic Acids Res. 12, 9383-9394, 1984
A;Title: Molecular cloning and carboxyl-propeptide analysis of human type III procollagen
A;Reference number: A93551; MUID:85087944; PMID:6096827
A;Accession: A93551
A;Molecule type: mRNA
A;Residues: 1065-1155, 'P', 1157-1466 <LOI>
A;Cross-references: UNIPARC:UPI0000173B88; EMBL:X01655; EMBL:X01742; NID:G29584; PIDN:CM
R;Miskulin, M.; Dalgleish, R.; Kluge-Beckerman, B.; Rennard, S.I.; Tolstoshev, P.; Brant
Biochemistry 25, 1408-1413, 1986
A;Title: Human type III collagen gene expression is coordinately modulated with the type
A;Reference number: I52393; MUID:86187804; PMID:3754462
A;Accession: I52393
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1161-1200 <MIS>
A;Cross-references: UNIPARC:UPI000016A685; GB:M13146; NID:G180415; PIDN:AAA52003.1; PID:
R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
A;Reference number: I59025; MUID:85216505; PMID:3958826
A;Accession: I59025
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1165-1196 <EMA>
A;Cross-references: UNIPARC:UPI000016A6B6; GB:M11134; NID:G180417; PIDN:AAA52004.1; PID:
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
J. Biol. Chem. 260, 4357-4363, 1985
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. P. F
A;Reference number: A92516; MUID:85157600; PMID:2579949
A;Accession: A92516
A;Molecule type: DNA
A;Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
A;Cross-references: UNIPARC:UPI0000173B89; GB:M10615; GB:M10793; GB:M10794; GB:M10795; G
A;Experimental source: liver
A;Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
ation
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (

3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C

C;Genetics: A;Gene: GDB:COL3A1
A;Cross-references: GDB:118729; OMIM:120180
A;Map position: 2q31-2q31
A;Introns: 27/1; 34/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
er of their length, is formed with desmosine cross-links made from lysine and allysine x
C;Function:
A;Description: structural component of extracellular fibrous polymer that maintains inte
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-153/Domain: amino-terminal propeptide #status predicted <PRO>
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>
F;154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
F;154-167/Region: collagen alpha 1(III) chain #status predicted
F;168-1196/Region: helical
F;1091-1093/Region: cell attachment (R-G-D) motif
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted
F;154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
F;161,1212/Modified site: allysine (Lys) #status predicted

Query Match 77.4%; Score 48; DB 1; Length 1466;
Best Local Similarity 80.0%; Pred. No. 5;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
Db 1170 GKGAGSGPG 1179
|||:|||||
|||:|||||

RESULT 11
JC4708
gelatin-binding 28K protein precursor - human
N;Alternate names: adipose specific collagen-like factor
C;Species: Homo sapiens (man)
C;Date: 10-May-1996 #sequence_revision 19-Jul-1996 #text_change 09-Jul-2004
C;Accession: JC4708; JG4944
R;Mada, K.; Okubo, K.; Shimomura, I.; Funahashi, T.; Matsuzawa, Y.; Matsubara, K.
Biochem. Biophys. Res. Commun. 221, 286-289, 1996
A;Title: cDNA cloning and expression of a novel adipose specific collagen-like factor, a
A;Reference number: JC4708; MUID:96224171; PMID:8619847
A;Accession: JC4708
A;Molecule type: mRNA
A;Residues: 1-244 <MAR>
A;Cross-references: UNIPROT:Q15848; UNIPARC:UPI0000034252; DDBJ:D45371; NID:9871886; PID
A;Experimental source: adipose tissue
R;Nakano, Y.; Tobe, T.; Choi-Miura, N.H.; Mazda, T.; Tomita, M.
J. Biochem. 120, 803-812, 1996
A;Title: Isolation and characterization of GBP28, a novel gelatin-binding protein purified
A;Reference number: JC4944; MUID:97103474; PMID:8947845
A;Accession: JC4944
A;Molecule type: protein
A;Residues: 19-38;93-100;101-112;135-149;173-178 <NAK>
A;Cross-references: UNIPARC:UPI000017A13F; UNIPARC:UPI000017A140; UNIPARC:UPI000017A141;
C;Comment: This protein is an endogenous factor that binds with a collagen-like domain.
C;Genetics:
A;Gene: apM1
C;Keywords: adipose tissue; glycoprotein; hydroxyproline
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-24/Product: gelatin-binding 28kDa protein #status experimental <MAT>
F;42-107/Region: collagen-like
F;114-241/Domain: complement C1q carboxyl-terminal homology <CIQ>
F;95/Modified site: 4-hydroxyproline (Pro) #status experimental
F;230/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 75.8%; Score 47; DB 2; Length 244;

Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 12
Db 63 GKGAGSGPGL 74
||||:|||||
||||:|||||

RESULT 12
S28774
collagen alpha chain - tube worm (Riftia pachyptila) (fragment)
C;Species: Riftia pachyptila
C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004
C;Accession: S28774; S22915; S17581
R;Mann, K.; Gaill, F.; Timpl, R.
Eur. J. Biochem. 210, 839-847, 1992
A;Title: Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen fro
A;Reference number: S28774; MUID:93130909; PMID:1483468
A;Accession: S28774
A;Molecule type: protein
A;Residues: 1-95; 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,
A;Cross-references: UNIPROT:P30754; UNIPARC:UPI000017A179
A;Note: we have shown the unidentified residues as Lys forming glycosylated 5-hydroxyls
R;Mann, K.; Gaill, F.; Timpl, R.
submitted to the Protein Sequence Database, July 1992
A;Description: Amino acid sequence and cell adhesion activity of a fibril-forming collag
A;Reference number: S22915
A;Accession: S22915
A;Molecule type: protein
A;Residues: 1-95, 'X', 97-107, 'X', 109-191, 'X', 193-260, 'X', 262-278, 'X', 280-572, 'X', 574-611,
A;Cross-references: UNIPARC:UPI000017A179
A;Note: 903-proline modified to 4-hydroxyproline was also found
R;Gaill, F.; Wiedemann, H.; Mann, K.; Kuehn, K.; Timpl, R.; Engel, J.
J. Mol. Biol. 221, 209-223, 1991
A;Title: Molecular characterization of cuticle and interstitial collagens from worms col
A;Reference number: S17581; MUID:92015209; PMID:1920405
A;Accession: S17581
A;Molecule type: protein
A;Residues: 8-45;525-545, 'X', 547-566, 'X', 568-572, 'X', 574-611, 'X', 613-618, 'X', 811-882 <GA
A;Cross-references: UNIPARC:UPI000017A17A; UNIPARC:UPI000017A17B; UNIPARC:UPI000017A17C
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (h
n 3- and 4-hydroxylated in the X-position. Lysines are 5-hydroxylated and subsequently a
C;Complex: homotrimer
C;Keywords: cell binding; coiled coil; extracellular matrix; glycoprotein; homotrimer; h
F;1-12/Domain: amino-terminal telopeptide (fragment) <NTE>
F;13-1023/Domain: collagenous #status experimental <COL>
F;1024-1027/Domain: carboxyl-terminal telopeptide (fragment) <CTE>
F;21,24,123,243,273,276,285,291,303,348,381,521,621,645/Modified site: 4-hydroxyproline (Pro
F;27,39,54,72,90,93,128,150,162,165,174,177,180,207,216,219,228,237,249,255,306,312,321,
711,714,717,723,744,759,774,783,792,816,843,849,855,861,867,888,894,915,945,963,966
F;53,161,165,416,551,647,773,815,1010,1013,1016,1019/Modified site: 3-hydroxyproline (Pr
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Modified site: 5-hydroxylysine (Lys)
F;96,108,192,261,279,573,612,657,738,765,810,927,936/Binding site: carboxylate (Lys) (c
F;183,342,546,567,939/Modified site: 5-hydroxylysine (Lys) #status experimental
F;351,933/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental

Query Match 75.8%; Score 47; DB 2; Length 1027;
Best Local Similarity 75.0%; Pred. No. 5.3;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 12
Db 736 GKGAGSGPGL 747
||||:|||||
||||:|||||

RESULT 13
I48103
type VII collagen - Chinese hamster (fragment)
C;Species: Cricetulus griseus (Chinese hamster)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: I48103
R;Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993

A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
A;Reference number: I48103; MUID:93271985; PMID:8499916
A;Accession: I48103
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-1549 <RES>
A;Cross-references: UNIPROT:Q60444; UNIPARC:UPI00000E753D; GB:L06863; NID:g388624; PIDN:
F:1484-1536/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Query Match 75.8%; Score 47; DB 2; Length 1549;
Best Local Similarity 80.0%; Pred. No. 7.9;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPG 10
||||| |||||
Db 242 GKGVEGNPG 251

RESULT 14
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N;Alternate names: procollagen alpha 5(IV) chain
N;Contains: collagen alpha 5(IV) chain precursor, leukocyte splice form
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence revision 27-Feb-1997 #text change 09-Jul-2004
C;Accession: S22917; A54365; A57079; A37122; I54317; A34850; S18850; I56971; I76598; A35
R;Zhou, J.; Hertz, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A;Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi
n Alport syndrome patient.
A;Reference number: S22917; MUID:92316923; PMID:1352287
A;Accession: S22917
A;Molecule type: mRNA
A;Residues: 1-967 <ZHO>
A;Cross-references: UNIPROT:P29400; UNIPARC:UPI0000173BDF; GB:M90464; NID:g180826; PIDN:
R;Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A;Title: Structure of the human type IV collagen COL4A5 gene.
A;Reference number: A54365; MUID:94165049; PMID:8120014
A;Accession: A54365
A;Molecule type: DNA
A;Residues: 1-922 <ZHA>
A;Cross-references: UNIPARC:UPI0000073BE0; GB:U04470; NID:g463378; GB:U04520; NID:g46342
R;Zhou, J.; Mochizuki, T.; Smets, H.; Antignac, C.; Laurila, P.; de Paeppe, A.; Tryggvason
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm
A;Reference number: A57079; MUID:93361972; PMID:8356449
A;Accession: A57079
A;Molecule type: DNA
A;Residues: 1-27 <ZHA>
A;Cross-references: UNIPARC:UPI000007378A; GB:Z37153; NID:g587203; PIDN:CAA85512.1; PID:
R;Pihlajaniemi, T.; Pohjola, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A;Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A;Reference number: A37122; MUID:90337990; PMID:2380186
A;Accession: A37122
A;Molecule type: mRNA
A;Residues: 84-433, 'GS', 442-624, 'LALQ', 629-666, 'FR', 669-887, 'R', 989-1264, 1271-1691 <PIH>
A;Cross-references: UNIPARC:UPI0000173BE1; GB:J05558; EMBL:M58526; NID:g1314209
A;Note: submitted to the EMBL Data Library, February 1991
A;Note: the authors translated the codon GCC for residue 115 as Val
R;Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
Hum. Mol. Genet. 1, 127-129, 1992
A;Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
A;Reference number: I54317; MUID:93244772; PMID:1363780
A;Accession: I54317
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 313-324, 'E', 326-330 <REN>
A;Cross-references: UNIPARC:UPI000016B3D0; GB:S59334; NID:g299946; PIDN:AAD13909.1; PID:
R;Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoeyhtyae, M.; Shows, T.B.; Tryggvason, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A;Title: Identification of a distinct type IV collagen alpha chain with restricted kidn

A;Reference number: A34850; MUID:90160375; PMID:1689491
A;Accession: A34850
A;Molecule type: mRNA
A;Residues: 914-1264, 1271-1691 <HOS>
A;Cross-references: UNIPARC:UPI000016A70B; EMBL:M31115; NID:g180824; PIDN:AAA52045.1; PI
R;Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
Genomics 9, 1-9, 1991
A;Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A;Reference number: A37969; MUID:91169491; PMID:2004755
A;Accession: S18850
A;Molecule type: DNA
A;Residues: 924-1264, 1271-1691 <ZH3>
A;Cross-references: UNIPARC:UPI000016A437; EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M6
EMBL:M63467; EMBL:M63468; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:g17792
R;Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; Ma
Kidney Int. 44, 1316-1321, 1993
A;Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex
A;Reference number: I56971; MUID:94133540; PMID:8301933
A;Accession: I56971
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1258-1276 <GUOI>
A;Cross-references: UNIPARC:UPI000016B421; GB:S69168; NID:g545095; PIDN:AAC60612.1; PID:
A;Note: kidney splice form
A;Accession: I76598
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1284-1291, 'TFGLGYLACL' <GUO2>
A;Cross-references: UNIPARC:UPI000011DDFD; GB:S69169; NID:g545097; PIDN:AAC60613.1; PID:
A;Note: frameshift mutation in patient with Alport syndrome
R;Myers, J.C.; Jones, T.A.; Pohjola, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.; SOJ
Am. J. Hum. Genet. 46, 1024-1033, 1990
A;Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the reg
A;Reference number: A35335; MUID:90252791; PMID:2339699
A;Accession: A35335
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1448-1477 <MYE>
A;Cross-references: UNIPARC:UPI0000173BE2
R;Nakazato, H.; Hattori, S.; Ushijima, T.; Matsuura, T.; Koitabashi, Y.; Takada, T.; Yosh
Kidney Int. 46, 1307-1314, 1994
A;Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in primord
A;Reference number: I56975; MUID:95156893; PMID:7853788
A;Accession: I56975
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1595-1602 <NAX>
A;Cross-references: UNIPARC:UPI00000004F8; GB:S75903; NID:g913882; PIDN:AAB33374.1; PID:
A;Note: premature termination mutation from a patient with Alport syndrome; one other mut
R;Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K.; H
Genomics 17, 485-489, 1993
A;Title: Identification of four novel mutations in the COL4A5 gene of patients with Alpo
A;Reference number: I54188; MUID:94010948; PMID:8406498
A;Accession: I54188
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1604-1607, 'VHDYKVC' <LEM>
A;Cross-references: UNIPARC:UPI000011F85C; GB:S65767; NID:g4255563; PIDN:AAD13967.1; PID:
A;Note: frameshift mutation from a patient with Alport syndrome; five other mutations ar
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (c
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A5; ATS
A;Cross-references: GDB:120596; OMIM:303630
A;Map position: Xq22-Xq22
A;Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3;
/3; 799/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C;Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 5
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
mer associations in the interrupted helical domain (with disulfide and desmosine cross-li
C;Function:
A;Description: minor structural component of extracellular basement membrane

C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <MAT1>
F;27-1264,1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #status predicted <MAT2>
F;27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F;42-1462/Region: interrupted helical
F;1463-1691/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F;1473-1573/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1593-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F;125/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;1482-1570,1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F;1527-1533,1638-1644/Disulfide bonds: #status predicted
F;1592-1684,1636-1687/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

Query Match 75.8%; Score 47; DB 1; Length 1691;
Best Local Similarity 72.7%; Pred. No. 8.6;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPGL 11
| | | | | | | | | |
Db 938 GEKSGKEPGL 948

RESULT 15
150696
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I50696
R;Nah, H.D.; Niu, Z.; Adams, S.L.
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not end
A;Reference number: A54041; MUID:94266842; PMID:8206952
A;Accession: I50696
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-310 <NAH>
A;Cross-references: UNIPROT:Q90612; UNIPARC:UPI00000FC2F0; EMBL:U07974; NID:G520456; PID
A;Gene: COL3A1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Query Match 74.2%; Score 46; DB 2; Length 310;
Best Local Similarity 80.0%; Pred. No. 2.4;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
| | | | | | | | | |
Db 152 GERGASGPG 161

RESULT 16
A55267
collagen alpha 5(IV) chain - dog (fragment)
C;Species: Canis lupus familiaris (dog)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 09-Jul-2004
C;Accession: A55267
R;Zheng, K.; Thorne, P.S.; Marrano, P.; Bauman, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A;Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X-li
en type IV.
A;Reference number: A55267; MUID:94224868; PMID:8171024
A;Accession: A55267
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-754 <ZHE>
A;Cross-references: UNIPROT:Q28247; UNIPARC:UPI0000126D56; GB:U07888; NID:G469547; PIDN:
C;Superfamily: collagen alpha 1(IV) chain

Query Match 74.2%; Score 46; DB 2; Length 754;
Best Local Similarity 72.7%; Pred. No. 5.8;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPGL 11
| | | | | | | | | |
Db 14 GEKSGKEPGL 24

RESULT 17
S31521
collagen COLF1 - freshwater sponge (Ephydatia muelleri)
C;Species: Ephydatia muelleri
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S31521
R;Exposito, J.Y.; van der Rest, M.; Garrone, R.
submitted to the EMBL Data Library, December 1992
A;Description: The complete intron/exon structure of E. muelleri collagen gene suggests
A;Reference number: S31521
A;Accession: S31521
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-812 <EXP>
A;Cross-references: UNIPROT:Q06452; UNIPARC:UPI000007D7E2; EMBL:X69818; NID:G429036; PID

Query Match 74.2%; Score 46; DB 2; Length 812;
Best Local Similarity 81.8%; Pred. No. 6.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPGL 11
| | | | | | | | | |
Db 81 GVGKAGSGPGL 91

RESULT 18
150694
collagen alpha 1(III) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I50694
R;Nah, H.D.; Niu, Z.; Adams, S.L.
J. Biol. Chem. 269, 16443-16448, 1994
A;Title: An alternative transcript of the chick type III collagen gene that does not end
A;Reference number: A54041; MUID:94266842; PMID:8206952
A;Accession: I50694
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-886 <NAH>
A;Cross-references: UNIPROT:P12105; UNIPARC:UPI000017125A; EMBL:U07973; NID:G520454; PID
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;30-90/Domain: von Willebrand factor type C repeat homology <VWC>

Query Match 74.2%; Score 46; DB 2; Length 886;
Best Local Similarity 80.0%; Pred. No. 6.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
| | | | | | | | | |
Db 728 GERGASGPG 737

RESULT 19
S28791
collagen alpha 1(XI) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C;Accession: S28791
R;Nah, H.D.; Barembaum, M.; Upholt, W.B.
J. Biol. Chem. 267, 22581-22586, 1992
A;Title: The chicken alpha1(XI) collagen gene is widely expressed in embryonic tissues.
A;Reference number: S28791; MUID:93054557; PMID:1429607
A;Accession: S28791
A;Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-888 <NAH>
A:Cross-references: UNIPROT:Q90796; UNIPARC:UPI000000FPAD3; EMBL:M88593; NID:g211619; PIDN:AA05773.1
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
F:665-887/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 74.2%; Score 46; DB 2; Length 888;
Best Local Similarity 80.0%; Pred. No. 6.8; Mismatches 1; Indels 0; Gaps 0;
Matches 8; Conservative 1;

Qy 1 GEKGAGSPG 10
|||||:|
Db 457 GEGAKGEPG 466

RESULT 20
A45748
collagen alpha 1(VII) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45748
R:Li, K.; Christiano, A.M.; Copeland, N.G.; Gilbert, D.J.; Chu, M.L.; Jenkins, N.A.; Uitto, J.; et al. Genomics 16, 733-739, 1993
A:Title: cDNA cloning and chromosomal mapping of the mouse type VII collagen gene (Col7a2)
A:Reference number: A45748; MUID:93315168; PMID:8325648
A:Accession: A45748
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-920 <LIH>
A:Cross-references: UNIPROT:Q63870; UNIPARC:UPI0000170C67; GB:S63654; NID:g386656; PIDN:AA05773.1
A:Experimental source: epidermal keratinocyte
A>Note: sequence extracted from NCBI backbone (NCBIN:135000, NCBIPI:135001)

Query Match 74.2%; Score 46; DB 2; Length 920;
Best Local Similarity 72.7%; Pred. No. 7; Mismatches 2; Indels 1; Gaps 0;
Matches 8; Conservative 2;

Qy 1 GEKGAGSPGL 11
|||||:|
Db 731 GEGAGNPGL 741

RESULT 21
S18251
collagen alpha 1(XI) chain - bovine (fragments)
C:Species: Bos primigenius taurus (cattle)
C>Date: 22-Nov-1993 #sequence_revision 23-Mar-1995 #text_change 09-Jul-2004
C:Accession: S18251; C46662; A56978; S65864; D46662; F46662; G46662; H46662; I46662; J46662; K46662; L46662; M46662; N46662; O46662; P46662; Q46662; R46662; S46662; T46662; U46662; V46662; W46662; X46662; Y46662; Z46662
R:Brown, K.E.; Lawrence, R.; Sonenshein, G.E. J. Biol. Chem. 266, 23268-23273, 1991
A:Title: Concerted modulation of alpha-1(XI) and alpha-2(V) collagen mRNAs in bovine vasculature
A:Reference number: S18251; MUID:92078200; PMID:1744123
A:Accession: S18251
A:Molecule type: mRNA
A:Residues: 1-911 <BRO>
A:Cross-references: UNIPROT:Q28083; UNIPARC:UPI0000126D29
R:Mayne, R.; Brewton, R.G.; Mayne, P.M.; Baker, J.R. J. Biol. Chem. 268, 9381-9386, 1993
A:Title: Isolation and characterization of the chains of type V/type XI collagen present in bovine cornea
A:Reference number: A46662; MUID:93252802; PMID:8486632
A:Accession: C46662
A:Molecule type: protein
A:Residues: 347-354; 356-363; 586-600; 912-924; 925-961; 962-998; 999-1024 <MAY>
A:Cross-references: UNIPARC:UPI000017739D; UNIPARC:UPI000017739E; UNIPARC:UPI000017739F; UNIPARC:UPI000017739G
A:Experimental source: vitreous humor
A>Note: sequence modified after extraction from NCBI backbone
R:Wu, J.J.; Eyre, D.R. J. Biol. Chem. 270, 18865-18870, 1995
A:Title: Structural analysis of cross-linking domains in cartilage type XI collagen. In: The Molecular Biology of the Extracellular Matrix, 2nd ed. (eds. J. Kleinman, J. Martin, J. VandeWoude), pp. 1-10. Academic Press, 1995
A:Reference number: A56978; MUID:95370194; PMID:7642541
A:Accession: A56978
A>Status: preliminary
A:Molecule type: protein

A:Residues: 206-229 <WUA>
A:Cross-references: UNIPARC:UPI00001773A4
R:Niyibizi, C.; Eyre, D.R. Eur. J. Biochem. 224, 943-950, 1994
A:Title: Structural characteristics of cross-linking sites in type V collagen of bone. C1
A:Reference number: S48210; MUID:95010086; PMID:7925418
A:Accession: S65864
A:Molecule type: protein
A:Residues: X', 273-298 <NIY>
A:Cross-references: UNIPARC:UPI00001773A5
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: hydroxyproline
F:211,223/Modified site: 4-hydroxyproline (Pro) #status experimental

Query Match 74.2%; Score 46; DB 2; Length 1024;
Best Local Similarity 72.7%; Pred. No. 7.8; Mismatches 1; Indels 0; Gaps 0;
Matches 8; Conservative 1;

Qy 1 GEKGAGSPGL 11
|||||:|
Db 479 GEGPQKPG 489

RESULT 22
T45467
collagen alpha 1(II) chain precursor [imported] - horse
N:Alternate names: type II collagen
C:Species: Equus caballus (domestic horse)
C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Jul-2004
C:Accession: T45467
R:Richardson, D.W.; Dodge, G.R. Submitted to the EMBL Data Library, June 1996
A:Description: Cloning of equine type II collagen and modulation of its expression in equine cells
A:Reference number: Z22977
A:Accession: T45467
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1418 <RIC>
A:Cross-references: UNIPROT:Q28396; UNIPARC:UPI000008834A; EMBL:U62528; PIDN:AA05773.1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match 74.2%; Score 46; DB 2; Length 1418;
Best Local Similarity 80.0%; Pred. No. 11; Mismatches 1; Indels 0; Gaps 0;
Matches 8; Conservative 1;

Qy 1 GEKGAGSPG 10
|||||:|
Db 702 GEGPQKPG 711

RESULT 23
A41182
collagen alpha 1(II) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 13-Aug-1999
C:Accession: A41182; A44885
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E. J. Biol. Chem. 266, 16862-16869, 1991
A:Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and expression
A:Reference number: A41182; MUID:91358489; PMID:1885613
A:Accession: A41182
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-1419 <MET>
A:Cross-references: UNIPARC:UPI0000177381; GB:M65161
R:Cheah, K.S.; Lau, E.T.; Au, P.K.; Tam, P.P. Development 111, 945-953, 1991
A:Title: Expression of the mouse alpha 1(II) collagen gene is not restricted to cartilage
A:Reference number: A44885; MUID:91347939; PMID:1879363
A:Accession: A44885
A:Molecule type: DNA
A:Residues: 1-28 <CHE>
A:Cross-references: UNIPARC:UPI00000004E7; GB:S63190; NID:g234368; PIDN:AA05773.1; PIDN:AA05773.1

A>Note: sequence extracted from NCBI backbone (NCBIN:63190, NCBI:63192)
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer
F:1191-1419/Domain: fibrillar collagen carboxyl-terminal homology <CC>
Query Match 74.2%; Score 46; DB 2; Length 1419;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GEKGAGSPG 10
||||| |||||
Db 703 GEKGEGAPG 712
RESULT 24
CGHUC
collagen alpha 1(II) chain precursor [validated] - human
N:Alternate names: procollagen alpha 1(II) chain
N:Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collagen
C:Species: Homo sapiens (man)
C:Date: 28-May-1986 #sequence revision 01-Sep-1995 #text change 31-Dec-2004
C:Accession: A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S637250; 137251; 137252; 137253; 137254; 155338; 155353; 161910
R:Ryan, M.C.; Sieraski, M.; Sandell, L.J.
Genomics 8, 41-48, 1990
A:Title: The human type II procollagen gene: identification of an additional protein-cod
A:Reference number: A38513; MUID:91184811; PMID:2081599
A:Accession: A38513
A:Molecule type: DNA
A:Residues: 1-103 <RYA>
A:Cross-references: UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP
g180884
R:Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.
Nucleic Acids Res. 17, 9473, 1989
A:Title: Nucleotide sequence of the full length cDNA encoding for human type II procolla
A:Reference number: S06715; MUID:90067946; PMID:2587267
A:Accession: S06715
A:Molecule type: mRNA
A:Residues: 1-28, 'R', 99-1487 <SU2>
A:Cross-references: UNIPARC:UPI0000126D15; EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID
A>Note: alternative splice form 1
R:Viikula, M.; Metsaeranta, M.; Syvaenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen, L.
Biochem. J. 285, 287-294, 1992
A:Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A:Reference number: S24270; MUID:92344585; PMID:1637314
A:Accession: S24270
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-28 <VIK>
A:Cross-references: UNIPARC:UPI0000173B64; EMBL:X58709; GB:S40537; NID:G35659
A>Note: this translation is not annotated in GenBank entry HSPROCOE1, release 111.0
R:Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.
Gene 44, 11-16, 1986
A:Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.
A:Reference number: A24828; MUID:87031574; PMID:3021582
A:Accession: A24828
A:Molecule type: DNA
A:Residues: 1-8, 'T', 10-28 <NUN>
A:Cross-references: UNIPARC:UPI000016A71A; GB:M25698; NID:g180872; PIDN:AAA52051.1; PID
R:Balwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A:Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A:Reference number: S06496; MUID:90026318; PMID:2803268
A:Accession: S06496
A:Molecule type: mRNA
A:Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834, 'F'
A:Cross-references: UNIPARC:UPI0000173B65; EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID
A>Note: alternative splice form 1
R:Ryan, M.C.; Sandell, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A:Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A:Reference number: A35428; MUID:90285153; PMID:2355003
A:Accession: A35428

A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 27-81, 'L', 83-103 <RYA2>
A:Cross-references: UNIPARC:UPI0000173B66
A>Note: alternative splice form 2; splicing appears to be under developmental regulation
R:Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A:Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A:Reference number: A30147; MUID:89233138; PMID:2714801
A:Accession: A30147
A:Molecule type: DNA
A:Residues: 104-157, 'P', 159-236 <SUM>
A:Cross-references: UNIPARC:UPI0000173B67; GB:J03065; GB:M23660; GB:M25655; GB:M25656; G
R:Ala-Kokko, L.; Balwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A:Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A:Reference number: A94227; MUID:90370826; PMID:1975693
A:Accession: A33116
A:Molecule type: DNA
A:Residues: 171-172, 'C', 174-175 <ALA>
A:Cross-references: UNIPARC:UPI0000173B68
A>Note: mutant sequence from a family with family with primary generalized osteoarthritis
R:Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A:Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A:Reference number: S64673; MUID:96195147; PMID:8660302
A:Accession: S64674
A:Molecule type: protein
A:Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DIA>
A:Cross-references: UNIPARC:UPI0000173B69; UNIPARC:UPI0000173B6A
R:Franc, S.; Warzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A:Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car
A:Reference number: S63514; MUID:96096730; PMID:8529631
A:Accession: S63514
A:Molecule type: protein
A:Residues: 243-261; 575-590; 756-763, 'X', 765-779 <FRA>
A:Cross-references: UNIPARC:UPI0000173B6B; UNIPARC:UPI0000173B6C; UNIPARC:UPI0000173B6D
R:Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A:Title: An RNA-splicing mutation (G+51VS20) in the type II collagen gene (COL2A1) in a
A:Reference number: 138867; MUID:95150028; PMID:7847372
A:Accession: 138867
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 440, 'G', 442-456, 'E', 458-480, 'P', 482-509 <TILL>
A:Cross-references: UNIPARC:UPI000006F3AF; EMBL:U15195; NID:g557053; PIDN:AA60370.1; PI
R:Ramirez, F.
submitted to the EMBL Data Library, December 1988
A:Reference number: S04892
A:Accession: S04892
A:Molecule type: mRNA
A:Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <RAM>
A:Cross-references: UNIPARC:UPI000016A700; EMBL:X13783; NID:g30037; PIDN:CAA32030.1; PID
R:Viikula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A:Title: Structural analyses of the polymorphic area in type II collagen gene.
A:Reference number: S05000; MUID:89325561; PMID:2753125
A:Accession: S05000
A:Molecule type: DNA
A:Residues: 630-640, 'A', 642-785 <VIK2>
A:Cross-references: UNIPARC:UPI0000173B6E; EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID
A34282.1; PID:g1335023; PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R:Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, I
J. Biol. Chem. 267, 22522-22526, 1992
A:Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A:Reference number: A44309; MUID:93054548; PMID:1429602
A:Accession: A44309
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA; mRNA
A:Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',
A:Cross-references: UNIPARC:UPI0000173B6F; GB:L00977; NID:g180812; PIDN:AAB23914.1; PID
A>Note: sequence extracted from NCBI backbone (NCBI:117273); parts of this sequence we

A;Note: this translation is not annotated and this publication is not cited in GenBank e
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individual
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184, 'GPSGKDGANGIPGP', 1185-1199 <TIL2>
A;Cross-references: UNIPARC:UPI000011F72; EMBL:M37126; NID:g180808; PIDN:AAAS2037.1; PID:
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>
A;Cross-references: UNIPARC:UPI000016A6B; GB:J00116; NID:g180395; PIDN:AAAS1997.1; PID:
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA, mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: UNIPARC:UPI000016A71B; EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XS', 1244-1246, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408 <VAN>
A;Cross-references: UNIPARC:UPI0000173B70; UNIPARC:UPI0000173B71; UNIPARC:UPI0000173B72
A;Note: chondrocalcin identified as released collagen I(II) chain carboxyl-terminal prope
R;Strom, C.M.; Upholt, W.B.
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human typ
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STR1>
A;Cross-references: UNIPARC:UPI000016A61A; EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:C
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909, 'PE', <STR2>
A;Cross-references: UNIPARC:UPI000006EA4D; GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:9
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro
gene
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>
A;Cross-references: UNIPARC:UPI0000173B73; GB:M12048; NID:g180017
A;Note: this translation is not annotated in GenBank entry HUMCCT2A, release 111.0
A;Note: the codons given for 1333-Lys (AGG) and 1350-Gly (GCA) are inconsistent with the
R;Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramirez,
Nucleic Acids Res. 13, 2207-2225, 1985
A;Title: Isolation and partial characterization of the entire human pro alpha 1(II) coll
A;Reference number: I37249; MUID:85215609; PMID:2987845
A;Accession: S59491
A;Molecule type: DNA
A;Residues: 7-28, 'R', 99-114, 541-578; 786-802; 1055-1056, 'N', 1058-1068, 'T', 1070-1109; 1200-1
A;Cross-references: UNIPARC:UPI0000071780; UNIPARC:UPI0000173B74; UNIPARC:UPI0000173B75;
A;Accession: I84453
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 7-28 <SAN2>
A;Cross-references: UNIPARC:UPI0000071780; UNIPARC:UPI0000173B74; UNIPARC:UPI0000173B75;

24938; NID:g30104
A;Note: the GenBank PID is based on an incorrect reading frame
Query Match 74.2%; Score 46; DB 1; Length 1487;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GKGAGSGSPG 10
DB 771 GKGPGGAPG 780
RESULT 25
B41182
collagen alpha 1(II) chain precursor (long splice form) - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 31-Dec-2004
C;Accession: B41182
R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
J. Biol. Chem. 266, 16862-16869, 1991
A;Title: Mouse type II collagen gene. Complete nucleotide sequence, exon structure, and
A;Reference number: A41182; MUID:91358489; PMID:1885613
A;Accession: B41182
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-1487 <MET>
A;Cross-references: UNIPROT:Q62031; UNIPROT:Q62032; UNIPROT:Q62033; UNIPARC:UPI0000177371
C;Superfamily: fibrillar collagen carboxyl-terminal homology; von Willebrand factor type
C;Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein; trimer;
F;33-91/Domain: von Willebrand factor type C repeat homology <VMC>
F;1259-1487/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 74.2%; Score 46; DB 2; Length 1487;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GKGAGSGSPG 10
DB 771 GKGPGGAPG 780
RESULT 26
A40333
collagen alpha 1'(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C;Accession: A40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Accession: A40333
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1492 <SUA>
A;Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; UNIPARC:UPI0000177382; GB:M63596
A;Note: this sequence is presented as substitutions relative to another sequence in a fi
es they replace; the appropriate interpretation of the sequence figure was reconstructed
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 74.2%; Score 46; DB 2; Length 1492;
Best Local Similarity 80.0%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GKGAGSGSPG 10
DB 776 GKGPGGAPG 785
RESULT 27

CGH202
collagen alpha 2(XI) chain precursor - human (fragment)
N:Alternate names: procollagen alpha 2(XI) chain
N:Contains: proline/arginine-rich protein (PARP)
C:Species: Homo sapiens (man)
C>Date: 07-Jun-1990 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999
C:Accession: S34790; PMID:8325374
R:Zhidkova, N.I.; Brewton, R.G.; Mayne, R.
FEBS Lett. 326, 25-28, 1993
A:Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage chain.
A:Reference number: S34790; MUID:93314796; PMID:8325374
A:Accession: S34790
A:Molecule type: mRNA
A:Residues: 1-663 <ZHI>
A:Cross-references: UNIPARC:UPI000016A416; EMBL:L119887; NID:G306439; PIDN:AAA35498.1; PI
R:Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.; J. Biol. Chem. 264, 13910-13916, 1989
A:Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and gene
A:Reference number: A32645; MUID:89340485; PMID:2760050
A:Accession: A32645
A:Molecule type: DNA; mRNA
A:Residues: 586-1546 <KIM>
A:Cross-references: UNIPARC:UPI000016A6F5; GB:J04974; NID:G180714; PIDN:AAA52034.1; PID:
A:Note: Parts of this sequence were determined by protein sequencing
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB: COL11A2
A:Cross-references: GDB:119788; OMIM:120290
A:Map position: 6p21.3-6p21.3
A:Introns: 1302/3; 1320/3; 1332/3; 1350/3; 1440/1; 1477/3
A:Note: The list of introns is incomplete
C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CGH
3(XI) chain (see PIR:CGH6C), initially linked by disulfide bonds among their carboxyl-
rmed with desmosine cross-links made from lysine and allysine residues
C:Function:
A:Description: structural component of extracellular fibrous polymer associated with cell
A:Note: may play a role in controlling the lateral growth of collagen II fibrils
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-254/Domain: non-collagenous (fragment) #status predicted <NC3>
F:1-187/Product: proline/arginine-rich PARP protein (fragment) #status predicted <PARP>
F:306-342/Domain: collagenous, triple helix #status predicted <COL2>
F:343-431/Region: non-collagenous #status predicted <NC2>
F:429-431/Region: cell attachment (R-G-D) motif
F:447-449/Region: cell attachment (R-G-D) motif
F:1257-1259/Region: cell attachment (R-G-D) motif
F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide
F:1381-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status atypical
F:109-163,1511-1545/Disulfide bonds: #status predicted
F:319/Modified site: allysine (Lys) #status predicted
F:426,1266/Modified site: 5-hydroxylysine (Lys) #status predicted
F:426,1266/Binding site: carboxylate (Lys) (covalent) #status predicted
F:927,933,1008,1017,1035,1038,1290,1296,1305,1317,1320/Modified site: 4-hydroxyproline
F:929/Modified site: 4-hydroxyproline (Pro) #status atypical
F:942,1023,1299/Modified site: 5-hydroxylysine (Lys) #status experimental
F:942,1023,1299/Binding site: carboxylate (Lys) (covalent) #status experimental
F:1427,1433,1450,1459/Disulfide bonds: interchain #status predicted
F:1460/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 74.2%; Score 46; DB 1; Length 1546;

Best Local Similarity 72.7%; Pred. No. 12;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GSKGEGSPGL 11

DB 526 GSKGEGSPGL 536

RESULT 28

CGH54B

collagen alpha 1(IV) chain precursor - mouse

C:Species: Mus musculus (house mouse)

C>Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004

C:Accession: A33525; S01454; A28666; A02864; A25636; A29301; S19079; A32003; A31766; S19

R:Muthukumar, G.; Blumberg, B.; Kurkinen, M.

J. Biol. Chem. 264, 6310-6317, 1989

A:Title: The complete primary structure for the alpha-1-chain of mouse collagen IV. Diff

A:Reference number: A33525; MUID:89197932; PMID:2703490

A:Accession: A33525

A:Molecule type: mRNA

A:Residues: 1-1669 <MUT>

A:Cross-references: UNIPARC:UPI000002724F; EMBL:J04694; NID:G556296; PID:

R:Wood, L.; Theriault, N.; Vogeli, G.

FEBS Lett. 227, 5-8, 1988

A:Title: cDNA clones completing the nucleotide and derived amino acid sequence of the al

A:Reference number: S01454; MUID:88112221; PMID:3338568

A:Accession: S01454

A:Molecule type: mRNA

A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71

R:Killen, P.D.; Burbelo, P.; Sakurai, Y.; Yamada, Y.

J. Biol. Chem. 263, 8706-8709, 1988

A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain

A:Reference number: A28066; MUID:88243724; PMID:3379041

A:Accession: A28066

A:Molecule type: mRNA

A:Residues: 1-129 <K11>

A:Cross-references: UNIPARC:UPI000016CCB0; EMBL:J03758; NID:G192669; PIDN:AAA37439.1; PI

R:Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Voss,

Eur. J. Biochem. 147, 217-224, 1985

A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1

A:Reference number: A02864; MUID:85127033; PMID:2578961

A:Accession: A02864

A:Molecule type: mRNA

A:Residues: 1276-1669 <OBE>

A:Cross-references: UNIPARC:UPI000016CC4B; EMBL:X02201; NID:G50233; PIDN:CAA26132.1; PID

R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.

Gene 43, 301-304, 1986

A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox

A:Reference number: A25636; MUID:86301886; PMID:3755692

A:Accession: A25636

A:Molecule type: mRNA

A:Residues: 1149-1396, 'S', 1398-1424 <NAT>

A:Cross-references: UNIPARC:UPI000016CC50; EMBL:M14042; NID:G192886; PIDN:AAA37342.1; PI

A:Note: The authors translated the codon CAG for residue 1374 as Arg

R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj

J. Biol. Chem. 262, 8496-8499, 1987

A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)

A:Reference number: A94680; MUID:87250460; PMID:3597383

A:Accession: A29301

A:Molecule type: mRNA

A:Residues: 1441-1669 <KUR>

A:Cross-references: UNIPARC:UPI000016CC4A; EMBL:M15832; NID:G192282; PIDN:AAA37340.1; PI

R:Killen, P.D.; Burbelo, P.D.; Martin, G.R.; Yamada, Y.

J. Biol. Chem. 263, 12310-12314, 1988

A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequen

A:Reference number: S19079; MUID:88315019; PMID:2842328

A:Accession: S19079

A:Molecule type: DNA

A:Residues: 1-28 <K12>

A:Cross-references: UNIPARC:UPI000000038F; EMBL:J03944; NID:G192673; PIDN:AAA37442.1; PI

R:Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.

J. Biol. Chem. 263, 19274-19277, 1988

A:Title: Head-to-head arrangement of murine type IV collagen genes.

A:Reference number: A92702; MUID:89066738; PMID:3198626

A:Accession: A32003

A:Molecule type: DNA

A:Residues: 1-28 <KAY>

A:Cross-references: UNIPARC:UPI000000038F; EMBL:J04448; NID:G192666; PIDN:AAA37437.1; PI

R:Burbelo, P.D.; Martin, G.R.; Yamada, Y.

Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988

A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promo

A;Reference number: A94220; MUID:89071759; PMID:3200851
A;Accession: A31766
A;Molecule type: DNA
A;Residues: 1-28 <BUR>
R;Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986
A;Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
A;Reference number: S19094; MUID:86196099; PMID:3009468
A;Accession: S19094
A;Molecule type: DNA
A;Residues: 1110-1133; 1189-1316; 1342-1383; 1418-1487 <SAK>
A;Cross-references: UNIPARC:UPI000016CC51; UNIPARC:UPI000016CC52; UNIPARC:UPI000016CC53;
R;Schuppan, D.; Timpl, R.; Glanville, R.W.
FEBS Lett. 115, 297-300, 1980
A;Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
A;Reference number: S16909; MUID:80246483; PMID:6772473
A;Accession: S16909
A;Molecule type: protein
A;Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-123
A;Cross-references: UNIPARC:UPI0000173BD2; UNIPARC:UPI0000173BD3; UNIPARC:UPI0000173BD4
R;Schuppan, D.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A;Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial amin
A;Reference number: A25991; MUID:82186723; PMID:6804236
A;Accession: A25991
A;Molecule type: protein
A;Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X', 100
61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-1119,
A;Cross-references: UNIPARC:UPI0000173BD5
A;Accession: B25991
A;Molecule type: protein
A;Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'Q', 1207, 'XE', 1210-1234,
3, 'SP', 1266, 'T', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1329
A;Cross-references: UNIPARC:UPI0000173BD5
R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A;Title: Subunit structure and assembly of the globular domain of basement-membrane coll
A;Reference number: S17801; MUID:84132058; PMID:6698021
A;Accession: S17801
A;Molecule type: protein
A;Residues: 1435-1443 <WEB>
A;Cross-references: UNIPARC:UPI0000173BDE
C;Genetics:
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
A;Note: the list of introns may be incomplete
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular mat
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;28-162/Domain: 7S <7SD>
F;163-1440/Domain: collagenous, triple helix <COL>
F;597-599/Region: cell attachment (R-G-D) motif
F;781-783/Region: cell attachment (R-G-D) motif
F;917-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;1441-1669/Domain: carboxyl-terminal nonhelical, NCI <NCI>
F;1441-1552/Region: duplication
F;1553-1669/Region: duplication
F;313639,41,434,467,470/Disulfide bonds: interchain #status predicted
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;971,974,977,986,989,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,1090,1
92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxypro
F;1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1505-1511,1616-1622/Disulfide bonds: #status predicted
Query Match 74.2%; Score 46; DB 1; Length 1669;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 1 GEKAGSGPL 11
|||||
Db 685 GERGADGLPL 695
|||||
RESULT 30
CGHULE
collagen alpha 1(XI) chain precursor - human
N;Alternate names: procollagen alpha 1(XI) chain
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 03-Oct-1995 #text_change 08-May-1998
C;Accession: A35239; A31795

Db 293 GEKGERGSPGI 303

RESULT 29

S40991
collagen alpha 1(IV) chain precursor - Caenorhabditis elegans
N;Alternate names: protein K04H4.1
C;Species: Caenorhabditis elegans
C;Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
C;Accession: S40991; S44442; S13651; B34476
R;Ainscough, R.
Submitted to the EMBL Data Library, October 1993
A;Reference number: S40991
A;Accession: S40991
A;Molecule type: DNA
A;Residues: 1-1744 <AIN>
A;Cross-references: UNIPROT:P17139; UNIPARC:UPI00001773C9; EMBL:Z27078; NID:9414627; PID
R;Kramer, J.M.
Submitted to the EMBL Data Library, December 1990
A;Reference number: S44442
A;Accession: S44442
A;Molecule type: DNA
A;Residues: 1-129, 'G'FGPMGLAGPPGQNGNFCRPLSGPPGEGGVNSQGRKGVKGESGRSGVPGLP', 209-281, 'P'
15, 'D', 817-1260, 'P', 1262-1707, 'P', 1709-1744 <KRA>
A;Cross-references: UNIPARC:UPI0000126D1B; EMBL:X56979; NID:66675; PIDN:CAA40299.1; PID:
R;Guo, X.; Johnson, J.J.; Kramer, J.M.
Nature 349, 707-709, 1991
A;Title: Embryonic lethality caused by mutations in basement membrane collagen of C. ele
A;Reference number: S13651; MUID:91141582; PMID:1996137
A;Accession: S13651
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-129, 'G'FGPMGLAGPPGQNGNFCRPLSGPPGEGGVNSQGRKGVKGESGRSGVPGLP', 209-281, 'P'
15, 'D', 817-1260, 'P', 1262-1515 <GUI>
A;Cross-references: UNIPARC:UPI00001773CA; EMBL:X56979
R;Guo, X.; Kramer, J.M.
J. Biol. Chem. 264, 17574-17582, 1989
A;Title: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are 1
A;Reference number: A34476; MUID:90008929; PMID:2793871
A;Accession: B34476
A;Molecule type: DNA
A;Residues: 1432-1499, 'Q', 1501-1707, 'P', 1709-1744 <GU2>
A;Cross-references: UNIPARC:UPI000016B8EF; EMBL:J05067; NID:6156255; PIDN:AA859179.1; PI
C;Genetics:
A;Gene: c1b-2; emb-9
A;Map position: 3
A;Introns: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;43-1515/Domain: collagenous, triple helix #status predicted <COL>
F;93-95/Region: cell attachment (R-G-D) motif
F;1053-1055/Region: cell attachment (R-G-D) motif
F;1396-1398/Region: cell attachment (R-G-D) motif
F;1516-1744/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
F;1516-1627, 1628-1744/Region: duplication
F;1580-1586, 1691-1697/Disulfide bonds: #status predicted
Query Match 74.2%; Score 46; DB 2; Length 1744;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GEKAGSGPL 11

Db 685 GERGADGLPL 695

RESULT 30

CGHULE
collagen alpha 1(XI) chain precursor - human
N;Alternate names: procollagen alpha 1(XI) chain
C;Species: Homo sapiens (man)
C;Date: 31-Mar-1990 #sequence_revision 03-Oct-1995 #text_change 08-May-1998
C;Accession: A35239; A31795

Mon Mar 13 10:53:09 2006

R; Yoshioka, H.; Ramirez, F.
 J. Biol. Chem. 265, 6423-6426, 1990
 A:Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and express
 A:Reference number: A35239; MUID:90202924; PMID:1690726
 A:Accession: A35239
 A:Molecule type: mRNA
 A:Residues: 1-558 <YOS>
 A:Cross-references: UNIPARC:UPI0000173BBF; GB:J05407
 R; Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;
 J. Biol. Chem. 263, 17159-17166, 1988
 A:Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type X
 cartilaginous tissue.
 A:Reference number: A92689; MUID:89034222; PMID:3182841
 A:Accession: A31795
 A:Molecule type: DNA
 A:Residues: 538-1806 <BER>
 A:Cross-references: UNIPARC:UPI0000173BC0; GB:J04177
 A:Note: parts of this sequence were determined by protein sequencing
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C:Genetics:
 A:Gene: GDB:COL1A1; COL1A6
 A:Cross-references: GDB:120595; OMIM:120280
 A:Map position: 1p21-1p21
 A:Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3
 A:Note: the list of introns is incomplete
 C:Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha
 3(XI) chain (see PIR:CGHU6C), initially linked by disulfide bonds among their carboxyl-
 rmed with desmosine cross-links made from lysine and allysine residues
 C:Function:
 A:Description: structural component of extracellular fibrous polymer associated with cel
 A:Note: may play a role in controlling the lateral growth of collagen II fibrils
 C:Superfamily: collagen alpha 1(IV) chain; fibrillar collagen carboxyl-terminal homology
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
 F:1-36/Domain: signal sequence #status predicted <SIG>
 F:13-260/Domain: PAPP-like #status predicted <PAPP>
 F:37-511/Domain: amino-terminal propeptide #status predicted <PRO>
 F:512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>
 F:512-527/Region: amino-terminal nonhelical telopeptide
 F:528-1542/Region: helical
 F:1543-1565/Region: carboxyl-terminal nonhelical telopeptide
 F:1566-1806/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:1583-1805/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:612-243.182-236/Disulfide bonds: #status predicted
 F:505/Modified site: allysine (Lys) #status predicted
 F:612.1452/Modified site: 5-hydroxylysine (Lys) #status predicted
 F:612.1452/Binding site: carbohydrate (Lys) (covalent) #status predicted
 Query Match 74.2%; Score 46; DB 1; Length 1806;
 Best Local Similarity 72.7%; Pred. No. 14;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAEGSPGL 11
 ||||| : |||
 Db 712 GEKGPQKPEGL 722
 RESULT 31
 I49558
 collagen alpha 1(I) precursor - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
 A:Accession: I49558
 R;Monson, J.M.; McCarthy, B.J.
 DNA 1, 59-69, 1981
 A:Title: Identification of a Balb/c mouse pro-alpha-1(I) procollagen gene: Evidence for
 A:Reference number: I49558; MUID:83157109; PMID:6219867
 A:Accession: I49558
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-198 <RES>
 A:Cross-references: UNIPARC:UPI000016CC45; GB:K03036; NID:gl92258; PIDN:AAA37332.1; PID:
 C:Genetics:

A:Gene: COL1A1
 A:Introns: 367/3; 54/3; 72/3; 108/3; 126/3; 144/3; 162/3
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 Query Match 72.6%; Score 45; DB 2; Length 198;
 Best Local Similarity 80.0%; Pred. No. 2.3;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GEKGAEGSPG 10
 ||||| : |||
 Db 4 GPKGADGSPG 13
 RESULT 32
 T20177
 hypothetical protein C53B4.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T20177
 R;Berks, M.
 submitted to the EMBL Data Library, December 1995
 A:Reference number: Z19233
 A:Accession: T20177
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-289 <WIL>
 A:Cross-references: UNIPROT:Q18799; UNIPARC:UPI0000080F8D; EMBL:Z68215; PIDN:CAA92453.1;
 A:Experimental source: clone C53B4
 C:Genetics:
 A:Gene: CESP:C53B4.5
 A:Map position: 4
 Query Match 72.6%; Score 45; DB 2; Length 289;
 Best Local Similarity 80.0%; Pred. No. 3.3;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAEGSPG 10
 ||||| : |||
 Db 236 GEKGANGPEP 245
 RESULT 33
 S57246
 ventral nervous system defective protein - fruit fly (Drosophila melanogaster)
 N:Alternate names: NK-2 homeotic protein
 C:Species: Drosophila melanogaster
 C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 05-Oct-2004
 A:Accession: S57246; B33976
 R;Jimenez, F.; Martin-Morris, L.E.; Velasco, L.; Chu, H.; Sierra, J.; Rosen, D.R.; White
 EMBO J. 14, 3487-3495, 1995
 A:Title: vnd, a gene required for early neurogenesis of Drosophila, encodes a homeodoma
 A:Reference number: S57246; MUID:95354667; PMID:7628450
 A:Accession: S57246
 A:Molecule type: mRNA
 A:Residues: 1-722 <JIM>
 A:Cross-references: UNIPROT:P22808; UNIPARC:UPI000012CA4F; EMBL:X87141; NID:gl045047; PI
 R;Kim, Y.; Nirenberg, M.
 Proc. Natl. Acad. Sci. U.S.A. 86, 7716-7720, 1989
 A:Title: Drosophila NK-homeobox genes
 A:Reference number: A33976; MUID:90046666; PMID:2573058
 A:Accession: B33976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 476-631, 'VG' <KIM>
 A:Cross-references: UNIPARC:UPI000016BC38; GB:M27290; NID:gl57635; PIDN:AAA28617.1; PID
 C:Genetics:
 A:Gene: FlyBase:vnd.
 A:Cross-references: FlyBase:FBgn0003986
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
 F:545-601/Domain: homeobox homology <HOX>
 Query Match 72.6%; Score 45; DB 2; Length 722;
 Best Local Similarity 81.8%; Pred. No. 8.2;

A;Note: submitted to the EMBL/GenBank/DBSJ databases by Prockop, D.J., 13-JUN-1988
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;
Nature 310, 337-340, 1984
A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of
A;Reference number: A93335; MUID:84270697; PMID:6462220
A;Accession: A93335
A;Molecule type: DNA
A;Residues: 1-58, 'Q', 60-181 <CHU>
A;Cross-references: UNIPARC:UPI00000173B3C; EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID
R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.
J. Biol. Chem. 262, 15151-15157, 1987
A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene enh
A;Reference number: 155254; MUID:88033098; PMID:2822714
A;Accession: 155254
A;Status: translation not shown; translated from GB/EMBL/DBSJ
A;Molecule type: DNA
A;Residues: 1-45 <ROS>
A;Cross-references: UNIPARC:UPI000016A6B2; GB:J02829; NID:g180387; PIDN:AAAS1993.1; PID:
R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.
Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987
A;Title: Regulatory elements in the first intron contribute to transcriptional control c
A;Reference number: A39943; MUID:88097389; PMID:3480516
A;Accession: A39943
A;Molecule type: DNA
A;Residues: 1-34 <BOR>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:J03559; NID:g180876; PIDN:AAAS2052.1; PID:
R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.
J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter s
A;Reference number: 155237; MUID:85130970; PMID:2857713
A;Accession: 155237
A;Status: translation not shown; translated from GB/EMBL/DBSJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:M10627; NID:g180383; PIDN:AAAS1992.1; PID:
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollis
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha1(I) amino-termina
some type VII
A;Reference number: A35233; MUID:90202908; PMID:2318855
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
A;Cross-references: UNIPARC:UPI0000173B3D
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R;Well, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
A;Reference number: S09400; MUID:89356643; PMID:2767050
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
A;Cross-references: UNIPARC:UPI0000173B3E
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha1
A;Reference number: A90567; MUID:71038625; PMID:5529814
A;Contents: CNB0-1, CNB2, CNB3, CNB4, CNB5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A;Cross-references: UNIPARC:UPI0000173B3F
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Notbohm, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, F.
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11372; MUID:90382436; PMID:2169412
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187; 274-287, 'P', 289 <BAE>
A;Cross-references: UNIPARC:UPI0000173B40; UNIPARC:UPI0000173B41
A;Note: sequence of collagen alpha 1(S) (I) isolated from bone after pepsin digestion

R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A;Reference number: 155342; MUID:92042092; PMID:1718984
A;Accession: 155342
A;Status: translated from GB/EMBL/DBSJ
A;Molecule type: mRNA
A;Residues: 258-268; 1347-1357 <DEA>
A;Cross-references: UNIPARC:UPI0000173B42; UNIPARC:UPI0000173B43; GB:S67495; NID:g239007
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508; PMID:4319110
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Cross-references: UNIPARC:UPI000014DF11
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the CD
A;Reference number: S15989; MUID:90326017; PMID:2374517
A;Accession: S15989
A;Molecule type: mRNA
A;Residues: 281-302; 402-420; 823-843; 925-944; 1026-1045; 1143-1162 <LAB>
A;Cross-references: UNIPARC:UPI0000173B44; UNIPARC:UPI0000173B45; UNIPARC:UPI0000173B46;
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N.
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of
A;Reference number: 152905; MUID:9339042; PMID:8339541
A;Accession: 152905
A;Status: translated from GB/EMBL/DBSJ
A;Molecule type: mRNA
A;Residues: 342-352, 'C', 354-359 <W12>
A;Cross-references: UNIPARC:UPI0000070EE6; GB:S64717; NID:g408195; PIDN:AA827677.1; PID:
A;Note: mutant sequence from patient with osteogenesis imperfecta
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha1
A;Reference number: A90476; MUID:84080385; PMID:6689127
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A;Cross-references: UNIPARC:UPI0000173B48; GB:K01228; NID:g180391; PIDN:AAA51995.1; PID:
A;Note: sequence partially completed for missing nucleotides by A29439
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multixon deletion in an osteogenesis imperfecta variant with increased type II I
A;Reference number: A22161; MUID:85104934; PMID:2981843
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594, 'R', 596-607 <CH3>
A;Cross-references: UNIPARC:UPI000011F796; GB:K03178; NID:g179612; NID:g179612; NID:g17961
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A;Reference number: A35336; MUID:90252792; PMID:2339700
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
A;Cross-references: UNIPARC:UPI0000173B4B
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolezzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the c
A;Reference number: 154365; MUID:95187161; PMID:7881420
A;Accession: 154365
A;Status: translated from GB/EMBL/DBSJ

A;Molecule type: DNA
A;Residues: 746-766, 'S', 768-781 <FOR>
A;Cross-references: UNIPARC:UPI000016A6FA; GB:L47667; NID:g1009093; PIDN:AAB59576.1; PID
R;Chesler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18223, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A;Reference number: A47426; MUID:93352646; PMID:8349697
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A;Cross-references: UNIPARC:UPI0000073A2A; GB:S64596; NID:g407589; PIDN:AAB27856.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIN:136444, NCHIP:136445)
A;Note: does not represent an experimentally determined sequence but three different mut
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Cross-references: UNIPARC:UPI0000173B4C
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276, 'H', 1278-1464 <CH5>
A;Cross-references: UNIPARC:UPI0000173B4D
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336, 1339-1464 <CH6>
A;Cross-references: UNIPARC:UPI0000173B4E
A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387, 'R', 1389-1464 <CH7>
A;Cross-references: UNIPARC:UPI0000173B4F
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nid
J. Biol. Chem. 263, 14605-14607, 1988

Query Match 72.6%; Score 45; DB 1; Length 1464;
Best Local Similarity 80.0%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GEKGAGSGP 10
| |||: |||
Db 749 GPKGAGSGP 758

RESULT 37
CGHU4B
collagen alpha 1(IV) chain precursor - human
N;Alternate names: procollagen alpha 1(IV) chain
C;Species: Homo sapiens (man)
C;Date: 28-May-1986 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C;Accession: S16876; A32117; S02738; S00048; S25826; A23115; S00207; S39614; A02863; A58
R;Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 264, 13565-13571, 1989
A;Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
A;Reference number: S16876; MUID:89340433; PMID:2701944
A;Accession: S16876
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1669 <SO11P>
A;Cross-references: UNIPROT:P02462; UNIPARC:UPI000004981D; EMBL:J04217; GB:J05039; NID:g
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
R;Soininen, R.; Huotari, M.; Hostalka, S.L.; Prockop, D.J.; Tryggvason, K.
J. Biol. Chem. 263, 17217-17220, 1988
A;Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen are
A;Reference number: A92690; MUID:89034231; PMID:3182844
A;Accession: A32117
A;Molecule type: DNA
A;Residues: 1-28 <SO12>
A;Cross-references: UNIPARC:UPI0000173BC1; EMBL:J04217; NID:g180759; PIDN:AAA53097.1; PI
R;Peschl, E.; Pollner, R.; Kuehn, K.
EMBO J. 7, 2687-2695, 1988
A;Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane c

A;Reference number: S02738; MUID:89030632; PMID:2846280
A;Accession: S02738
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-6, 'L', 8-28 <POE>
A;Cross-references: UNIPARC:UPI0000173BC2; EMBL:X12784; NID:g30072
R;Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, R.;
Eur. J. Biochem. 168, 529-536, 1987
A;Title: Completion of the amino acid sequence of the alpha1 chain of human basement mem
A;Reference number: S00048; MUID:88029471; PMID:3311751
A;Accession: S00048
A;Molecule type: mRNA
A;Residues: 1-318, 'A', 320-944 <BRAL>
A;Cross-references: UNIPARC:UPI000016A708; EMBL:X05561; NID:g30066; PIDN:CAA29075.1; PID
A;Accession: S25826
A;Molecule type: protein
A;Residues: 271-318, 'A', 320-554 <BRA2>
A;Cross-references: UNIPARC:UPI0000173BC3
R;Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
Eur. J. Biochem. 152, 213-219, 1985
A;Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (7S c
A;Reference number: A23115; MUID:86004708; PMID:4043082
A;Accession: A23115
A;Molecule type: protein
A;Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>
A;Cross-references: UNIPARC:UPI0000173BC4
A;Experimental source: placenta
A;Note: the amino end of the mature form is blocked
R;Soininen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.
FEBS Lett. 225, 188-194, 1987
A;Title: Complete primary structure of the alpha(1)-chain of human basement membrane (ty
A;Reference number: S00207; MUID:88083584; PMID:3691802
A;Accession: S00207
A;Molecule type: mRNA
A;Residues: 244-530 <SO13>
A;Cross-references: UNIPARC:UPI0000173BC5; EMBL:Y00706; NID:g29548; PIDN:CAA68698.1; PID
R;Edle, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
EMBO J. 12, 4795-4802, 1993
A;Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen m
A;Reference number: S39614; MUID:94038963; PMID:8223488
A;Accession: S39614
A;Molecule type: protein
A;Residues: 371-554 <EBL>
A;Cross-references: UNIPARC:UPI0000173BC6
R;Babel, W.; Glanville, R.W.
Eur. J. Biochem. 143, 545-556, 1984
A;Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid seq
A;Reference number: A02863; MUID:85003629; PMID:6434307
A;Accession: A02863
A;Molecule type: protein
A;Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-1
A;Cross-references: UNIPARC:UPI0000173BC7
A;Experimental source: placenta
R;Glanville, R.W.; Rauter, A.
Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
A;Title: Pepsin fragments of human placental basement-membrane collagens showing interrup
A;Reference number: S16908; MUID:82005835; PMID:6792033
A;Accession: A58517
A;Molecule type: protein
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553; 1389-1405, 'XX', 1408-1409, 'X', 1411-14
A;Cross-references: UNIPARC:UPI0000173BC8; UNIPARC:UPI0000173BC9
R;MacWright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
Biochemistry 22, 4940-4948, 1983
A;Title: Isolation and characterization of pepsin-solubilized human basement membrane (ty
A;Reference number: S16910; MUID:84053346; PMID:6416291
A;Accession: S16910
A;Molecule type: protein
A;Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549; 939-940, 'M', 942-944, 'V', 946, 'X', 948-
A;Cross-references: UNIPARC:UPI0000173BCA; UNIPARC:UPI0000173BCB
A;Experimental source: placenta
R;Pihlajaniemi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.; P
J. Biol. Chem. 260, 7681-7687, 1985
A;Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen r

A;Reference number: S01466; MUID:85207819; PMID:2581969
A;Accession: S01466
A;Molecule type: mRNA
A;Residues: 1256-1669 <PTH>
A;Cross-references: UNIPARC:UPI000016A6B8; EMBL:M10940; NID:G180421; PIDN:AAAS2006.1; PI
R;Brinker, J.M.; Gudas, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kafalides, N.A.;
Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A;Title: Restricted homology between human alpha-1 type IV and other procollagen chains.
A;Reference number: S16879; MUID:85216555; PMID:2582422
A;Accession: S16879
A;Molecule type: mRNA
A;Residues: 1259-1669 <BRI>
A;Cross-references: UNIPARC:UPI000016A707; EMBL:M11315; NID:G180817; PIDN:AAAS2042.1; PI
R;Oberbauer, I.; Laurent, M.; Schwart, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A;Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A;Reference number: A02864; MUID:85127033; PMID:2578961
A;Accession: S19091
A;Molecule type: protein
A;Residues: 1435-1461, 'H', 1463-1482, 'X', 1484-1491, 1501-1514, 'X', 1516-1519, 1534-1553, 'X',
A;Cross-references: UNIPARC:UPI0000173BCC; UNIPARC:UPI0000173BCE; UNIPARC:UPI0000173BCE;
R;Siebold, B.; Deutzmann, R.; Kuehn, K.
Eur. J. Biochem. 176, 617-624, 1988
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterm
A;Reference number: S02550; MUID:89005112; PMID:2844531
A;Contents: annotation; disulfide bonds
C;Genetics:
A;Gene: GDB:COL4A1
A;Cross-references: GDB:119791; OMIM:120130
A;Map position: 13q34-13q34
A;Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3; 156/3; 184/3; 205/3; 217/3; 231/
A;Exons: 782/1; 820/1; 876/1; 906/1; 957/1; 1020/1; 1066/3; 1109/1; 1136/1; 116
A;Complex: type IV collagen is a heterotrimer of two alpha 1(IV) chains and one alpha 2
A;Title: Associations among trimer amino-terminal domains (disulfide and desmosine cross-links), dim
A;Title: Associations in the interrupted helical domain (with disulfide and desmosine cr
C;Function:
A;Description: structural component of extracellular basement membrane
A;Superfamily: collagen alpha 1(IV) chain
A;Keywords: basement membrane; blocked amino end; cell binding; coiled coil; duplication
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1669/Product: collagen alpha 1(IV) chain #status predicted <NAT>
F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F;163-1440/Domain: interrupted helical <COL>
F;414-452/Region: integrin binding #status experimental
F;597-599/Region: cell attachment (R-G-D) motif
F;917-919/Region: cell attachment (R-G-D) motif
F;968-970/Region: cell attachment (R-G-D) motif
F;1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;17/Modified site: blocked amino end (Ala) (in mature form) #status experimental
F;31.36.39.41.125.434.467.470/Disulfide bonds: interchain #status predicted
F;45.48.78.90.129.156.172.217.228.231.277.295.298.322.343.361.460.463.497.527.540.543.57
F;1081.1084.1099.1117.1132.1150.1165.1182.1185.1188.1206.1235.1265.1283.1304.1319.1328.134
F;45.48.78.90.129.156.172.217.228.231.277.295.298.322.343.361.460.463.497.527.543.573.582.61
F;99.1117.1132.1150.1165.1182.1185.1188.1206.1235.1265.1283.1304.1319.1328.1340.1356.1371.
F;54.63.75.84.87.96.102.105.108.111.117.120.123.138.141.147.150.153.159.167.178.181.184.
F;419.422.425.439.445.448.451.479.485.491.494.503.512.518.524.530.546.549.552.555.561.567
F;145.748.751.754.763/Modified site: 4-hydroxyproline (Pro) #status experimental
F;126/Binding site: carboxylate (Asn) (covalent) #status experimental
F;129/Modified site: lysine (Lys) #status predicted
F;172.540.947/Modified site: 5-hydroxylysine (Lys) #status atypical
F;272.645.839/Modified site: 4-hydroxyproline (Pro) #status atypical
F;446-447/Cleavage site: Gly-Ile (Gelatinase B) #status predicted
F;766.775.784.787.790.796.799.804.810.816.822.834.860.863.869.872.875.887.890.893.899.90
F;1129.1138.1141.1159.1171.1176.1179.1194.1200.1203.1215.1224.1227.1244.1247.1250.1256.
F;1143/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1120.1268/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
F;1120.1268/Binding site: carboxylate (Lys) (covalent) (partial) #status experimental
F;1214.1424/Modified site: 3-hydroxyproline (Pro) #status absent
F;1392.1395.1398.1404/Modified site: 4-hydroxyproline (Pro) #status experimental
F;1460-1548.1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F;1505-1511.1616-1622/Disulfide bonds: #status predicted

F;1570-1662, 1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted
Query Match 72.6%; Score 45; DB 1; Length 1669;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GEKGAGSGPL 11
Db 679 GEKGAGSQPGI 689
RESULT 38
A45407
collagen alpha 3(IV) chain - sea urchin (Strongylocentrotus purpuratus)
C;Species: Strongylocentrotus purpuratus (purple urchin)
C;Date: 22-Sep-1993 #sequence revision 18-Nov-1994 #text_change 31-Dec-2004
C;Accession: A45407; A43903; A23940
R;Exposito, J.Y.; D'Alessio, M.; Di Liberto, M.; Ramirez, F.
J. Biol. Chem. 268, 5249-5254, 1993
A;Title: Complete primary structure of a sea urchin type IV collagen alpha chain and ana
A;Reference number: A45407; MUID:93186842; PMID:8444899
A;Accession: A45407
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1752 <EXP>
A;Cross-references: UNIPROT:Q26312; UNIPARC:UPI000001773DE
A;Note: sequence extracted from NCBI backbone (NCBIP:126841)
R;Weisel, G.M.; Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A;Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously produ
A;Reference number: A43903; MUID:92038439; PMID:1936564
A;Accession: A43903
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: P, 633-1537, 'G' <WES>
A;Cross-references: UNIPARC:UPI000007C802; GB:S64572; NID:G238616; PIDN:AAAB20270.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBIN:64572; NCBIP:64573)
R;Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A;Title: Structure and developmentally regulated expression of a Strongylocentrotus purp
A;Reference number: A23940; MUID:86205894; PMID:3458186
A;Accession: A23940
A;Molecule type: DNA
A;Residues: 742-812 <VEN>
A;Cross-references: UNIPARC:UPI00001773DF; EMBL:M13206
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;29-162/Domain: amino-terminal nonhelical, 7S <7SD>
F;162-1523/Region: interrupted helical
F;1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;129/Modified site: allysine (Lys) #status predicted
Query Match 72.6%; Score 45; DB 2; Length 1752;
Best Local Similarity 80.0%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GEKGAGSGPG 10
Db 777 GEKGAGQIPG 786
RESULT 39
A31893
collagen alpha 1(IV) chain precursor - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 21-May-1990 #sequence revision 21-May-1990 #text_change 09-Jul-2004
C;Accession: A31893; A26692; A19442; S00020
R;Blumberg, B.; Mackrell, A.J.; Fessler, J.H.
J. Biol. Chem. 263, 18328-18337, 1988
A;Title: Drosophila basement membrane procollagen alpha-1(IV). II. Complete cDNA sequen
A;Reference number: A31893; MUID:89054012; PMID:3142875
A;Accession: A31893

A;Molecule type: mRNA
A;Residues: 1-1775 <BLU>
A;Cross-references: UNIPROT:P08120; UNIPARC:UPI0000126D1C; EMBL:M23704; NID:G157029; PIDN:CAA23
R;Blumberg, B.; Mackrell, A.J.; Olson, P.F.; Kurkinen, M.; Monson, J.M.; Natzie, J.E.; F
J. Biol. Chem. 262, 5947-5950, 1987
A;Title: Basement membrane procollagen IV and its specialized carboxyl domain are conserved
A;Reference number: A26692; MUID:87194801; PMID:3106346
A;Accession: A26692
A;Molecule type: mRNA
A;Residues: 1065-1775 <BLU2>
A;Cross-references: UNIPARC:UPI00001773AF; EMBL:J02727
R;Monson, J.M.; Natzie, J.; Friedman, J.; McCarthy, B.J.
Proc. Natl. Acad. Sci. U.S.A. 79, 1761-1765, 1982
A;Title: Expression and novel structure of a collagen gene in Drosophila.
A;Reference number: A19442; MUID:82197577; PMID:6210912
A;Accession: A19442
A;Molecule type: DNA
A;Residues: 762-947, 'S', 949-996, 'T', 998-1230 <MON>
A;Cross-references: UNIPARC:UPI000016BB4B; GB:J01074; EMBL:V00200; NID:G7736; PIDN:CAA23
R;Cecchini, J.P.; Knibiehl, B.; Mirre, C.; le Parco, Y.
Eur. J. Biochem. 165, 587-593, 1987
A;Title: Evidence for a type-IV-related collagen in Drosophila melanogaster. Evolutionary
A;Reference number: S00020; MUID:87246644; PMID:3109906
A;Accession: S00020
A;Molecule type: DNA
A;Residues: 1355-1356, 'K', 1358-1359, 'K', 1361-1372, 'I', 1374-1495, 'R', 1497-1506, 'RA', 1509,
A;Cross-references: UNIPARC:UPI00001773B0; EMBL:M28334
C;Genetics:
A;Gene: FlyBase:Cg25C
A;Cross-references: FlyBase:FBgn0000299
A;Introns: 7/2, 23/3, 339/3, 505/2, 989/1, 1312/1, 1689/3
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-1775/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F;65-67/Region: cell attachment (R-G-D) motif
F;130-132/Region: cell attachment (R-G-D) motif
F;238-240/Region: cell attachment (R-G-D) motif
F;297-299/Region: cell attachment (R-G-D) motif
F;892-894/Region: cell attachment (R-G-D) motif
F;1075-1077/Region: cell attachment (R-G-D) motif
F;1173-1175/Region: cell attachment (R-G-D) motif
F;1225-1227/Region: cell attachment (R-G-D) motif
F;1545-1775/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>
F;1545-1655/Domain: repeat NCI #status predicted <NC11>
F;1656-1775/Domain: repeat NCI #status predicted <NC12>
F;72/Binding site: carbohydrate (Aen) (covalent) #status predicted
F;570,573/Disulfide bonds: interchain #status predicted
F;1611-1617,1720-1727/Disulfide bonds: #status predicted

Query Match 72.6%; Score 45; DB 2; Length 1775;
Best Local Similarity 72.7%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
|:|:| |
Db 131 GKGAGSGPGL 141

RESULT 40
T29773
hypothetical protein T21D9.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
A;Accession: T29773
R;Nhan, M.; Hawkins, J.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid T21D9.
A;Reference number: Z20683
A;Accession: T29773
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-396 <NHA>

A;Cross-references: UNIPROT:Q22651; UNIPARC:UPI0000076EBB; EMBL:U40937; PIDN:AAA81693.1;
C;Genetics:
A;Gene: CESP:T21D9.1
A;Introns: 77/1, 315/3
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8
Query Match 71.0%; Score 44; DB 2; Length 396;
Best Local Similarity 72.7%; Pred. No. 6.8;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
|:|:| |
Db 261 GKGAGSGPGL 271

RESULT 41
A27353
collagen alpha 1(III) chain precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 28-Aug-1989 #sequence_revision 30-Sep-1990 #text_change 09-Jul-2004
C;Accession: A27353; A22287; F50066
R;Wood, L.; Theriault, N.; Vogell, G.
Gene 61, 225-230, 1987
A;Title: Complete nucleotide sequence of the N-terminal domains of the murine alpha-1 ty
A;Reference number: A91588; MUID:88167858; PMID:3443309
A;Accession: A27353
A;Molecule type: mRNA
A;Residues: 1-488 <WOO>
A;Cross-references: UNIPROT:P08121; UNIPARC:UPI000016CC49; GB:M18933; NID:G192276; PIDN:J
A;Note: in the helical region, Gly-X-Y repeat is interrupted by 287-Val instead of Gly
R;Kilau, G.; Mudryj, M.; de Crombrughe, B.
J. Biol. Chem. 260, 3773-3777, 1985
A;Title: Identification of the promoter and first exon of the mouse alpha 1(III) collagen
A;Reference number: A92513; MUID:85131189; PMID:3972847
A;Accession: A22287
A;Molecule type: DNA
A;Residues: 1-28 <LJA>
A;Cross-references: UNIPARC:UPI0000177380
C;Genetics:
A;Introns: 29/1
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; sulfoprotein; trimer; triple
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-154/Domain: amino-terminal propeptide #status predicted <APR>
F;32-92/Domain: von Willebrand factor type C repeat homology <VWC>
F;155-488/Product: collagen alpha 1(III) chain (fragment) #status predicted <MAT>
F;164-488/Region: helical (fragment)
F;40,152,159/Binding site: sulfate (Tyr) (covalent) #status predicted

Query Match 71.0%; Score 44; DB 2; Length 488;
Best Local Similarity 72.7%; Pred. No. 8.3;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
|:|:| |
Db 272 GKGAGSGPGL 282

RESULT 42
CGBO6C
collagen alpha 1(II) chain precursor - bovine (tentative sequence) (fragments)
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 17-May-1996 #text_change 09-Jul-2004
C;Accession: A90369; A90396; A92210; S03940; A90189; A05039; A02859
R;Miller, E.J.; Lunde, L.G.
Biochemistry 12, 3153-3159, 1973
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
A;Reference number: A90369; MUID:73258693; PMID:4732855
A;Contents: composition of CNBr and CNBr4
A;Accession: A90369
A;Molecule type: protein
A;Residues: 1-15 <MIL>
A;Cross-references: UNIPROT:P02459; UNIPARC:UPI0000173B79

F;19-56/Domain: amino-terminal nonhelical #status predicted <NC2>
F;57-519/Region: interrupted helical
F;520-680/Domain: amino-terminal nonhelical #status predicted <NC1>
F;553-679/Domain: complement C1q carboxyl-terminal homology <C1Q>
F;617/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 71.0%; Score 44; DB 1; Length 680;
Best Local Similarity 72.7%; Pred. No. 12;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11

||||| |||:

Db 174 GEKGAGVPGM 184

RESULT 44

A53330

collagen alpha 2(IX) chain precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004

C;Accession: A53330; #22216

R;Peraelae, M.; Elima, K.; Metsaeranta, M.; Rosati, R.; de Crombrughe, B.; Vuorio, E.

J. Biol. Chem. 269, 5064-5071, 1994

A;Title: The exon structure of the mouse alpha2(IX) collagen gene shows unexpected diver

A;Reference number: A53330; MUID:94148964; PMID:8106484

A;Accession: A53330

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-688 <PER>

A;Cross-references: UNIPROT:Q07643; UNIPARC:UPI000002803D; GB:Z22923; NID:g311949; PIDN:

R;Elima, K.; Metsaeranta, M.; Kallio, J.; Peraelae, M.; Eroola, I.; Garofalo, S.; de Cro

Biochim. Biophys. Acta 1130, 78-80, 1992

A;Title: Specific hybridization probes for mouse alpha-2(IX) and alpha-1(X) collagen mRN

A;Reference number: S22215; MUID:92182017; PMID:1543751

A;Accession: S22216

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 521-524, 'M', 526-609, 'R', 611-665, 'G', 667-668 <ELI>

A;Cross-references: UNIPARC:UPI000016CBAC; EMBL:X63014; NID:g49810; PIDN:CAA44742.1; PID

A;Note: the authors translated the codon GAC for residue 526 as His and GGT for residue

C;Genetics:

A;Intons: 24/3; 49/3; 61/3; 82/3; 100/3; 112/3; 120/3; 138/3; 156/3; 172/3; 191/3; 209/

3; 534/1; 597/1; 623/1

C;Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; hydroxyllysine

Query Match 71.0%; Score 44; DB 2; Length 688;

Best Local Similarity 72.7%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11

||||| |||:

Db 453 GEKGSGEPGL 463

RESULT 45

A33526

collagen alpha 2(IV) chain precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004

C;Accession: A33526; A24432; D24432; A25066; B25066; A24364; S19081; B29301; A24628; B32

R;Saus, J.; Quinones, S.; Mackrell, A.; Blumberg, B.; Muthukumar, G.; Phlajantemi, T.

J. Biol. Chem. 264, 6318-6324, 1989

A;Title: The complete primary structure of mouse alpha-2(IV) collagen. Alignment with m

A;Reference number: A33526; MUID:89197933; PMID:2703491

A;Accession: A33526

A;Molecule type: mRNA

A;Residues: 1-1707 <SAU>

A;Cross-references: UNIPROT:P08122; UNIPARC:UPI00000295D8; EMBL:J04695; NID:g556298; PID

R;Kurkinen, M.; Bernard, M.P.; Barlow, D.P.; Chow, L.T.

Nature 317, 177-179, 1985

A;Title: Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha-2(IV)

A;Reference number: A93367; MUID:85296379; PMID:3839908

A;Accession: A24432

A;Molecule type: mRNA

A;Residues: 967-1096, 'G', 1098-1109 <KU1>

A;Cross-references: UNIPARC:UPI000016CC5A; EMBL:X02896; NID:g50263; PIDN:CAA26655.1; PID

A;Note: the authors translated the codon AAC for residue 964 as Lys

A;Accession: D24432

A;Molecule type: DNA

A;Residues: 964-1096, 'G', 1098-1109 <KU2>

A;Cross-references: UNIPARC:UPI00001773B1; EMBL:X02899

R;Schwarz, U.; Schuppan, D.; Oberbauer, I.; Glanville, R.W.; Deutzmann, R.; Timpl, R.;

Eur. J. Biochem. 157, 49-56, 1986

A;Title: Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-

A;Reference number: A25066; MUID:86220192; PMID:3011432

A;Accession: A25066

A;Molecule type: mRNA

A;Residues: 970-1480 <SC1>

A;Cross-references: UNIPARC:UPI00001773B2; EMBL:X04647

A;Accession: B25066

A;Molecule type: protein

A;Residues: 979-1058; 1065-1101; 1105-1222; 1226-1310; 1326-1335; 1351-1480 <SC2>

A;Cross-references: UNIPARC:UPI00001773B3; UNIPARC:UPI00001773B4; UNIPARC:UPI00001773B5;

R;Vogeli, G.; Horn, E.; Carter, J.; Kaytes, P.S.

FEBS Lett. 206, 29-32, 1986

A;Title: Proposed alignment of helical interruptions in the two subunits of the basement

A;Reference number: A24364; MUID:87005245; PMID:3758345

A;Accession: A24364

A;Molecule type: mRNA

A;Residues: 1041-1050, 'R', 1052-1170, 'S', 1172-1178, 'R', 1180-1240, 'E', 1242-1327, 'A', 1329-1

A;Cross-references: UNIPARC:UPI000016CC4E; EMBL:X04410; NID:g50240; PIDN:CAA27998.1; PID

R;Kaytes, P.S.; Theriault, N.Y.; Vogeli, G.

Gene 54, 141-146, 1987

A;Title: Homologies between the non-collagenous C-terminal (NC1) globular domains of the

A;Reference number: S19080; MUID:87277427; PMID:3609751

A;Accession: S19081

A;Molecule type: mRNA

A;Residues: 1466-1622, 'H', 1624-1707 <KA1>

A;Cross-references: UNIPARC:UPI00001773B9; GB:X04410; NID:g50240; PIDN:CAA27998.1; PID:G

R;Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj

J. Biol. Chem. 262, 8496-8499, 1987

A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)

A;Reference number: A94680; MUID:87250460; PMID:3597383

A;Accession: B29301

A;Molecule type: mRNA

A;Residues: 1481-1707 <KUR>

A;Cross-references: UNIPARC:UPI000016CC4C; EMBL:M15833; NID:g192284; PIDN:AAA37341.1; PI

R;Schwarz-Magdolen, U.; Oberbauer, I.; Kuehn, K.

FEBS Lett. 208, 203-207, 1986

A;Title: cDNA and protein sequence of the NC1 domain of the alpha-2-chain of collagen IV

A;Reference number: A24628; MUID:87054581; PMID:3780963

A;Accession: A24628

A;Molecule type: mRNA

A;Residues: 1480-1572, 'L', 1574-1622, 'H', 1624-1707 <SCH>

A;Cross-references: UNIPARC:UPI00001773BA; EMBL:X04647

R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.

J. Biol. Chem. 263, 19274-19277, 1988

A;Title: Head-to-head arrangement of murine type IV collagen genes.

A;Reference number: A92702; MUID:89066738; PMID:3198626

A;Accession: B32003

A;Molecule type: DNA

A;Residues: 1-33 <KA2>

A;Cross-references: UNIPARC:UPI000016CCAF; EMBL:J04448; NID:g192666; PIDN:AAA37438.1; PI

R;Burbelo, P.D.; Martin, G.R.; Yamada, Y.

Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988

A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional promo

A;Reference number: A94220; MUID:89071759; PMID:3200851

A;Accession: B31766

A;Molecule type: DNA

A;Residues: 1-60 <BUR>

A;Cross-references: UNIPARC:UPI000016CCB3; EMBL:M23333

R;Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.

Eur. J. Biochem. 139, 401-410, 1984

A;Title: Subunit structure and assembly of the globular domain of basement-membrane coll

A;Reference number: S17801; MUID:84132058; PMID:6698021

A;Accession: S19086

Mon Mar 13 10:53:09 2006

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z23034
 A;Accession: T46404
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-201 <AA>
 A;Cross-references: UNIPROT:Q9NT93; UNIPARC:UPI00000701E3; EMBL:AL137461
 A;Experimental source: adult testis; clone DKF2p434K0621
 C;Genetics:
 A;Note: DKF2p434K0621.1

Query Match 69.4%; Score 43; DB 2; Length 201;
 Best Local Similarity 72.7%; Pred. No. 5.1;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11
 ||||| |||||
 Db 130 GEKGRPGEPGL 140

RESULT 48
 A55047
 collagen alpha 1(V) - chicken (fragment)
 C;Species: Gallus gallus (chicken)
 C;Date: 08-Jul-1995 #sequence_revision 03-Aug-1995 #text_change 31-Dec-2004
 C;Accession: A55047
 R;Gordon, M.K.; Foley, J.W.; Birk, D.E.; Fitch, J.M.; Linsenmayer, T.F.
 J. Biol. Chem. 269, 24959-24966, 1994
 A;Title: Type V collagen and Bowman's membrane. Quantitation of mRNA in corneal epitheli
 A;Reference number: A55047; MUID:95014268; PMID:7929179
 A;Accession: A55047
 A;Status: preliminary
 A;Molecule type: mRNA
 A;Residues: 1-252 <GOR>
 A;Cross-references: UNIPROT:Q91014; UNIPROT:Q91014; UNIPARC:UPI000017739C; GB:L31917
 C;Superfamily: fibrillar collagen carboxyl-terminal homology

Query Match 69.4%; Score 43; DB 2; Length 252;
 Best Local Similarity 70.0%; Pred. No. 6.4;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSPG 10
 ||||| |||||
 Db 10 GDKGEGSPG 19

RESULT 49
 T20435
 hypothetical protein E03G2.3 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T20435
 R;McMurray, A.
 submitted to the EMBL Data Library, November 1995
 A;Reference number: Z19275
 A;Accession: T20435
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-274 <WLL>
 A;Cross-references: UNIPROT:Q27276; UNIPARC:UPI000017B9A4; EMBL:Z68113; PIDN:CAA92149.1;
 A;Experimental source: clone E03G2
 C;Genetics:
 A;Gene: CESP:E03G2.3
 A;Map position: X
 A;Introns: 42/3; 89/2; 197/3; 239/3

Query Match 69.4%; Score 43; DB 2; Length 274;
 Best Local Similarity 80.0%; Pred. No. 7;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSPG 10
 ||||| |||||
 Db 129 GEKGMEGLPG 138

A;Molecule type: protein
 A;Residues: 1475-1481, 'X', 1483-1487 <WEB>
 A;Cross-references: UNIPARC:UPI00001773BB
 C;Genetics:
 A;Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
 F;1-28/Domain: signal sequence #status predicted <SIG>
 F;29-1707/Product: collagen alpha 1(IV) chain #status predicted <MAT>
 F;29-171/Domain: 7S #status predicted <7SD>
 F;58-1480/Domain: collagenous #status predicted <COL>
 F;141-143/Region: cell attachment (R-G-D) motif
 F;360-362/Region: cell attachment (R-G-D) motif
 F;779-781/Region: cell attachment (R-G-D) motif
 F;884-886/Region: cell attachment (R-G-D) motif
 F;965-967/Region: cell attachment (R-G-D) motif
 F;1223-1225/Region: cell attachment (R-G-D) motif
 F;1447-1449/Region: cell attachment (R-G-D) motif
 F;1481-1707/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NC1>
 F;1481-1589/Domain: repeat NCI #status predicted <NC11>
 F;1590-1707/Domain: repeat NCI #status predicted <NC12>
 F;42.47.51.53.481.483/Disulfide bonds: interchain #status predicted
 F;138.1270/Binding site: carboxylate (Asn) (covalent) #status predicted
 F;656-676.154-1550.1653-1660/Disulfide bonds: #status predicted
 F;985.988.997.1003.1028.1031.1067.1101.1113.1119.1143.1170.1200.1242.1305.1368.1391/Modi
 F;985.988.997.1003.1028.1031.1067.1101.1113.1119.1143.1170.1200.1242.1305.1368.1391/Modi
 F;1009.1012.1018.1021.1024.1037.1040.1043.1046.1052.1058.1070.1098.1110.1128.1140.1149.1
 77.1333.1386.1401.1408.1420.1423.1429.1444.1465.1468.1471.1477/Modified site: hydroxypro

Query Match 71.0%; Score 44; DB 2; Length 1707;
 Best Local Similarity 72.7%; Pred. No. 29;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11
 ||||| |||||
 Db 1249 GERPGAGSPGL 1259

RESULT 46
 T24064
 hypothetical protein R09A8.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: T24064
 R;Wilkinson, J.
 submitted to the EMBL Data Library, November 1995
 A;Reference number: Z19836
 A;Accession: T24064
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-152 <WIL>
 A;Cross-references: UNIPROT:Q21855; UNIPARC:UPI000007928C; EMBL:Z68009; PIDN:CAA92006.1;
 A;Experimental source: clone R09A8
 C;Genetics:
 A;Gene: CESP:R09A8.4
 A;Map position: X

Query Match 69.4%; Score 43; DB 2; Length 152;
 Best Local Similarity 72.7%; Pred. No. 3.9;
 Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11
 ||||| |||||
 Db 73 GEKGPAGSPGL 83

RESULT 47
 T46404
 hypothetical protein DKF2p434K0621.1 - human (fragment)
 C;Species: Homo sapiens (man)
 C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C;Accession: T46404
 R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. .

RESULT 50

T32765

hypothetical protein F33D11.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004

C;Accession: T32765

R;Sammons, L.; Wohldmann, P.; Mullen, G.

submitted to the EMBL Data Library, December 1997

A;Description: The sequence of C. elegans cosmid F33D11.

A;Reference number: Z21222

A;Accession: T32765

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-281 <SAM>

A;Cross-references: UNIPROT:O44774; UNIPARC:UPI000007A941; EMBL:AF039720; PIDN:AAB96697.

A;Experimental source: strain Bristol N2; clone F33D11

C;Genetics:

A;Gene: CESP:F33D11.3

A;Map position: 1

A;Introns: 48/3; 102/1

Query Match

Best Local Similarity 69.4%; Score 43; DB 2; Length 281;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY

1 GEGKAGSGPG 10

||:|||||

Db 168 GEGSGEGPG 177

RESULT 51

C88638

protein F58F6.1 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C;Accession: C88638

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology

A;Reference number: A75000; MUID:99069613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and

A;Accession: C88638

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-281 <STO>

A;Cross-references: UNIPROT:O44174; UNIPARC:UPI0000082EA6; GB:chr_IV; PIDN:AAB88358.1; F

C;Genetics:

A;Gene: F58F6.1

A;Map position: 4

Query Match

Best Local Similarity 69.4%; Score 43; DB 2; Length 281;

Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY

1 GEGKAGSGPGL 11

||:|||||

Db 154 GDKGNDGAPGL 164

RESULT 52

A32249

collagen - sea urchin (Paracentrotus lividus) (fragment)

C;Species: Paracentrotus lividus [common urchin]

C;Date: 17-Aug-1989 #sequence_revision 17-Aug-1989 #text_change 09-Jul-2004

C;Accession: A32249

R;Saitta, B.; Buttice, G.; Gambino, R.

Biochem. Biophys. Res. Commun. 158, 633-639, 1989

A;Title: Isolation of a putative collagen-like gene from the sea urchin Paracentrotus li

A;Reference number: A32249; MUID:89149773; PMID:2537631

A;Accession: A32249

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA

A;Residues: 1-290 <SAI>

A;Cross-references: UNIPROT:Q26054; UNIPARC:UPI0000177399

C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix

Query Match

Best Local Similarity 69.4%; Score 43; DB 2; Length 290;

Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY

1 GEGKAGSGPG 10

||:|||||

Db 19 GEGKAGSGQG 28

RESULT 53

T21668

hypothetical protein F32G8.5 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T21668

R;McMurray, A.

submitted to the EMBL Data Library, May 1996

A;Reference number: Z19456

A;Accession: T21668

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-294 <WIL>

A;Cross-references: UNIPROT:Q19979; UNIPARC:UPI000007A0B1; EMBL:Z72509; PIDN:CAA96649.1;

A;Experimental source: clone F32G8

C;Genetics:

A;Gene: CESP:F32G8.5

A;Map position: 5

A;Introns: 47/3

Query Match

Best Local Similarity 69.4%; Score 43; DB 2; Length 294;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY

1 GEGKAGSGPG 10

||:|||||

Db 163 GENGADGNPG 172

RESULT 54

A44984

collagen - nematode (Haemonchus contortus)

C;Species: Haemonchus contortus

C;Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 15-Sep-2003

C;Accession: A44984

R;Shamansky, L.M.; Pratt, D.; Boisvenue, R.J.; Cox, G.N.

Mol. Biochem. Parasitol. 37, 73-86, 1989

A;Title: Cuticle collagen genes of Haemonchus contortus and Caenorhabditis elegans are h

A;Reference number: A44984; MUID:90136718; PMID:2615789

A;Accession: A44984

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-295 <SHA>

A;Cross-references: UNIPARC:UPI000017A127; GB:J04670

Query Match

Best Local Similarity 69.4%; Score 43; DB 2; Length 295;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY

1 GEGKAGSGPGL 11

||:|||||

Db 245 GESGSDGQPGI 255

RESULT 55

A61396

collagen alpha 1(II) chain - golden hamster (fragments)

C;Species: Mesocricetus auratus (golden hamster)
C;Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004
C;Accession: A61396
R;Cizdziel, P.E.; Hosoi, J.; Montgomery, J.C.; Wiseman, R.W.; Barrett, J.C.
Mol. Carcinog. 4, 14-24, 1991
A;Title: Loss of a tumor suppressor gene function is correlated with downregulation of c
A;Reference number: A61396; MUID:91182265; PMID:2009131
A;Accession: A61396
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-323 <CIZ>
A;Cross-references: UNIPROT:Q7M099; UNIPARC:UPI0000177383
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;120-323/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 69.4%; Score 43; DB 2; Length 323;
Best Local Similarity 80.0%; Pred. No. 8.2; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
Db 14 GEPGREGSPG 23

RESULT 56
T26281
hypothetical protein W08D2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T26281
R;Swinburne, J.; Ainscough, R.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z20188
A;Accession: T26281
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-358 <WIL>
A;Cross-references: UNIPROT:Q23222; UNIPARC:UPI000007AB54; EMBL:Z70271; PIDN:CAA94234.1;
A;Experimental source: clone W08D2
C;Genetics:
A;Gene: CESP:W08D2.6
A;Map position: 4
A;Introns: 57/3; 320/2

Query Match 69.4%; Score 43; DB 2; Length 358;
Best Local Similarity 70.0%; Pred. No. 9.1; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
Db 299 GEQGADGGPG 308

RESULT 57
A55797
collagen precursor, saccule-specific - bluegill
C;Species: Lepomis macrochirus (bluegill)
C;Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 15-Sep-2003
C;Accession: A55797
R;Davis, J.G.; Oberholtzer, J.C.; Burns, F.R.; Greene, M.I.
Science 267, 1031-1034, 1995
A;Title: Molecular cloning and characterization of an inner ear-specific structural pro
A;Reference number: A55797; MUID:95167486; PMID:7863331
A;Accession: A55797
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <DAV>
A;Cross-references: UNIPARC:UPI000014730F; GB:U17431; NID:9687605; PIDN:AAA69978.1; PID
F;58-271/Domain: collagenous, triple helix #status predicted <COL>
F;281-410/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 69.4%; Score 43; DB 2; Length 423;
Best Local Similarity 72.7%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
Db 197 GEXGERGPPGL 207

RESULT 58
S53787
collagen alpha chain - Paralvinella grasslei (fragments)
C;Species: Paralvinella grasslei
C;Date: 19-Mar-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C;Accession: S53787
R;Gall, F.; Mann, K.; Wiedemann, H.; Engel, J.; Timpl, R.
J. Mol. Biol. 246, 284-294, 1995
A;Title: Structural comparison of cuticle and interstitial collagens from annelids livin
A;Reference number: S53786; MUID:95173973; PMID:7869380
A;Accession: S53787
A;Molecule type: protein
A;Residues: 1-90;91-254;255-304;305-374;375-438 <GAI>
A;Cross-references: UNIPROT:Q7M3U7; UNIPARC:UPI000017A15F; UNIPARC:UPI000017A160; UNIPAR
C;Keywords: coiled coil; extracellular matrix; glycoprotein

Query Match 69.4%; Score 43; DB 2; Length 438;
Best Local Similarity 70.0%; Pred. No. 11; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
Db 2 GERGRDSPG 11

RESULT 59
T31631
hypothetical protein Y57A10A.i - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31631
R;Smye, R.
submitted to the EMBL Data Library, September 1999
A;Reference number: Z21048
A;Accession: T31631
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-458 <WIL>
A;Cross-references: UNIPROT:Q9NA83; UNIPARC:UPI0000082499; EMBL:AL117195; NID:el549729;
A;Experimental source: clone Y57A10A
C;Genetics:
A;Gene: CESP:Y57A10A.i
A;Introns: 8/3; 54/3; 112/3; 151/1

Query Match 69.4%; Score 43; DB 2; Length 458;
Best Local Similarity 63.6%; Pred. No. 12; Indels 0; Gaps 0;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
Db 408 GEPGADGPGM 418

RESULT 60
B31795
collagen alpha 1(XI) chain - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
C;Accession: B31795
R;Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;
J. Biol. Chem. 263, 17159-17166, 1988
A;Title: Cloning and sequencing of pro-alpha(XI) collagen cDNA demonstrates that type x
cartilaginous tissue.
A;Reference number: A92689; MUID:89034222; PMID:3182841
A;Accession: B31795
A;Molecule type: mRNA

A;Residues: 1-482 <BER>
A;Cross-references: UNIPROT:P20909; UNIPARC:UPI0000126D2B
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;259-481/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 69.4%; Score 43; DB 2; Length 482;
Best Local Similarity 80.0%; Pred. No. 12; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 2

Qy 1 GEKGAEPSG 10
||| ||||| ||
Db 57 GEKGAEPPG 66

RESULT 61
T33149
hypothetical protein F29C4.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Dec-2004
C;Accession: T33149
R;Bradshaw, H.; Clarke, K.
submitted to the EMBL Data Library, May 1998
A;Description: The sequence of C. elegans cosmid F29C4.
A;Reference number: 221291
A;Accession: T33149
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-614 <BRA>
A;Cross-references: UNIPROT:O76368; UNIPARC:UPI0000075CEE; EMBL:AF067616; PIDN:AAC19194.
A;Experimental source: strain Bristol N2; clone F29C4
C;Genetics:
A;Gene: CESP:F29C4.8
A;Map position: 4
A;Introns: 1/3; 32/1; 43/3; 62/3; 98/3; 110/3; 122/3; 239/3; 298/3; 350/3; 374/3; 427/3;
C;Superfamily: fibrillar collagen carboxyl-terminal homology; von Willebrand factor type

Query Match 69.4%; Score 43; DB 2; Length 614;
Best Local Similarity 72.7%; Pred. No. 16; Indels 0; Gaps 0;
Matches 8; Conservative 1; Mismatches 2

Qy 1 GEKGAEPSG 11
||||| : |||||
Db 564 GEKGQGIPL 574

RESULT 62
A45137
collagen alpha 4(IV) chain - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A45137
R;Kamagata, Y.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 267, 23753-23758, 1992
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of
A;Reference number: S28777; MUID:93054733; PMID:1429714
A;Accession: A45137
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-623 <KAM>
A;Cross-references: UNIPROT:P55787; UNIPARC:UPI0000126D55
A;Experimental source: basement membrane
A;Note: sequence extracted from NCBI backbone (NCBIP:118549)
C;Superfamily: collagen alpha 1(IV) chain

Query Match 69.4%; Score 43; DB 2; Length 623;
Best Local Similarity 80.0%; Pred. No. 16; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 2

Qy 1 GEKGAEPSG 10
||||| |||||
Db 276 GEKGLPGSPG 285

RESULT 63

SL3301

collagen alpha 1(X) chain precursor - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 21-Nov-1993 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004

C;Accession: S13301

R;Thomas, J.T.; Kwan, A.P.L.; Grant, M.E.; Boot-Handford, R.P.

Biochem. J. 273, 141-148, 1991

A;Title: Isolation of cDNAs encoding the complete sequence of bovine type X collagen. EV.

A;Reference number: S13301; MUID:91113131; PMID:1703407

A;Accession: S13301

A;Molecule type: mRNA

A;Residues: 1-674 <THO>

A;Cross-references: UNIPROT:P23206; UNIPARC:UPI0000126D26; EMBL:X53556; NID:q263; PIDN:CJ

C;Genetics:

A;Gene: COL10A1

C;Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

C;Keywords: coiled coil; extracellular matrix; glycoprotein; homotrimer

F;1-18/Domain: signal sequence #status predicted <SIG>

F;19-674/Product: collagen alpha 1(X) chain #status predicted <MAT>

F;547-673/Domain: complement C1q carboxyl-terminal homology <CIQ>

Query Match 69.4%; Score 43; DB 2; Length 674;
Best Local Similarity 72.7%; Pred. No. 17; Indels 0; Gaps 0;
Matches 8; Conservative 0; Mismatches 3Qy 1 GEKGAEPSGL 11
||||| |||||
Db 174 GEKGTSGVPGL 184

RESULT 64

S20819

collagen alpha 3(IX) chain precursor - chicken

C;Species: Gallus gallus (chicken)

C;Date: 22-Nov-1993 #sequence_revision 09-Mar-1996 #text_change 09-Jul-2004

C;Accession: S20819; S22429; S22918; S22338; C18856; S22241

R;Brewton, R.G.; Ouspenskaya, M.V.; van der Rest, M.; Mayne, R.

Eur. J. Biochem. 205, 443-449, 1992

A;Title: Cloning of the chicken alpha-3(IX) collagen chain completes the primary structure

A;Reference number: S20819; MUID:92241276; PMID:1572350

A;Accession: S20819

A;Molecule type: mRNA

A;Residues: 1-675 <ERE>

A;Cross-references: UNIPROT:Q90800; UNIPARC:UPI00000FDCD2; EMBL:X64712; NID:q63316; PIDN:

R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;

J. Biol. Chem. 267, 10070-10076, 1992

A;Title: Cloning and developmental expression of the alpha3 chain of chicken type IX col

A;Reference number: S22429; MUID:92250566; PMID:1577778

A;Accession: S22429

A;Molecule type: mRNA

A;Residues: 1-195, 'G', 197-675 <HA1>

A;Cross-references: UNIPARC:UPI000017A144; EMBL:M83179

A;Note: 353-Arg, 386-Leu and 548-Arg were also found

R;Har-El, R.; Sharma, Y.D.; Aguilera, A.; Ueyama, N.; Wu, J.J.; Eyre, D.R.; Juricic, L.;

submitted to the EMBL Data Library, February 1992

A;Description: Cloning and developmental expression of the alpha 3 chain of chicken type

A;Reference number: S22918

A;Accession: S22918

A;Molecule type: mRNA

A;Residues: 1-195, 'G', 197-405, 'S', 407-675 <HA2>

A;Cross-references: UNIPARC:UPI0000126D50; EMBL:M83179; NID:q211040; PIDN:AAB59960.1; PI:

R;Mayne, R.; van der Rest, M.; Ninomiya, Y.; Olsen, B.R.

Ann. N. Y. Acad. Sci. 460, 38-46, 1985

A;Title: The structure of type IX collagen.

A;Reference number: S22238; MUID:86185164; PMID:3868958

A;Accession: S22238

A;Molecule type: protein

A;Residues: 540-548 <MAY>

A;Cross-references: UNIPARC:UPI000017A145

R;Ninomiya, Y.; van der Rest, M.; Mayne, R.; Lozano, G.; Olsen, B.R.

Biochemistry 24, 4223-4229, 1985


```

F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-1142/Product: collagen alpha 1(XIX) chain #status predicted <MAT>
F;24-292/Domain: amino-terminal nonhelical #status predicted <NHI>
F;292-1123/Region: interrupted helical
F;1124-1142/Domain: carboxyl-terminal nonhelical #status predicted <NH2>

Query Match          69.4%; Score 43; DB 2; Length 1142;
Best Local Similarity 70.0%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSPG 10
    |||||:|
Db 798 GEGKSDGPPG 807

RESULT 67
A38587
collagen, cornea-specific - chicken
C;Species: Gallus gallus (chicken)
C;Date: 14-Feb-1992 #sequence_revision 15-Aug-1997 #text_change 09-Jul-2004
C;Accession: SI6501; A38587
R;Marchant, J.K.; Linsemayer, T.F.; Gordon, M.K.
Proc. Natl. Acad. Sci. U.S.A. 88, 1560-1564, 1991
A;Title: cDNA analysis predicts a cornea-specific collagen.
A;Reference number: A38587; MUID:91142213; PMID:1705041
A;Accession: SI6501
A;Molecule type: mRNA
A;Residues: 1-1146 <MAR>
A;Cross-references: UNIPROT:Q90584; UNIPARC:UPI00000FC061; EMBL:M60172; NID:G211609; PID
A;Accession: A38587
A;Molecule type: mRNA
A;Residues: 1-174, 'X', 176-233, 'X', 235-344, 'X', 346-408, 'X', 410-499, 'X', 501-876, 'X', 878-11
A;Cross-references: UNIPARC:UPI000017A187; GB:M60172
C;Keywords: cornea

Query Match          69.4%; Score 43; DB 2; Length 1146;
Best Local Similarity 80.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPG 10
    |||||:|
Db 281 GEGKATGPPG 290

RESULT 68
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: B40333
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.
J. Cell Biol. 115, 565-575, 1991
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A;Reference number: A40333; MUID:92011898; PMID:1918153
A;Accession: B40333
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1486 <SUA>
A;Cross-references: UNIPROT:Q91718; UNIPARC:UPI0000173B50; GB:M63595
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F;37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match          69.4%; Score 43; DB 1; Length 1486;
Best Local Similarity 72.7%; Pred. No. 37;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGL 11
    |||||:|
Db 620 GEGKLVGAPGL 630

```

RESULT 69

CGHU2V

collagen alpha 2 (V) chain precursor - human

C;Species: Homo sapiens (man)

C;Date: 31-Jul-1989 #sequence_revision 28-Jul-1995 #text_change 09-Jul-2004

C;Accession: A31427; A54555; S43643; A25874; I55239; I59025; A25374; A30017

R;Woodbury, D.; Benson-Chanda, V.; Ramirez, F.

J. Biol. Chem. 264, 2735-2738, 1989

A;Title: Amino-terminal propeptide of human pro-alpha2(V) collagen conforms to the struc

A;Reference number: A31427; MUID:89123368; PMID:2914927

A;Accession: A31427

A;Molecule type: mRNA

A;Residues: 1-463 <WOO>

A;Cross-references: UNIPROT:P05997; UNIPARC:UPI000016A628; GB:J04478; NID:G179697; PIDN:J

R;Greenspan, D.S.; Lee, S.T.; Lee, B.S.; Hoffman, G.G.

Gene Expr. 1, 29-39, 1991

A;Title: Homology between alpha2(V) and alpha1(III) collagen promoters and evidence for i

A;Reference number: A54555; MUID:92314691; PMID:1820205

A;Accession: A54555

A;Molecule type: DNA

A;Residues: 1-32 <GRE>

A;Cross-references: UNIPARC:UPI00000049F; GB:M58529; NID:G180834; PIDN:AAC41699.1; PID:

R;Moradi-Ameli, M.; Rousseau, J.C.; Kleman, J.P.; Champilaud, M.F.; Boutillon, M.M.; Ber

Eur. J. Biochem. 221, 987-995, 1994

A;Title: Diversity in the processing events at the N-terminus of type-V collagen.

A;Reference number: S43642; MUID:94237164; PMID:8181482

A;Accession: S43643

A;Molecule type: protein

A;Residues: 288-291, 'P', 293-294, 'X', 296-297, 606, 'X', 608-617 <MOR>

A;Cross-references: UNIPARC:UPI0000173B93; UNIPARC:UPI0000173B94

R;Weil, D.; Bernard, M.; Gargano, S.; Ramirez, F.

Nucleic Acids Res. 15, 181-198, 1987

A;Title: The pro alpha 2(V) collagen gene is evolutionarily related to the major fibrill

A;Reference number: A25874; MUID:87146331; PMID:3029669

A;Accession: A25874

A;Molecule type: mRNA; DNA

A;Residues: 398-1496 <WEI>

A;Cross-references: UNIPARC:UPI000016A633; GB:X04758; NID:G29588; PIDN:CAA28454.1; PID:G

A;Experimental source: rhabdomyosarcoma cell line

R;Myers, J.C.; Loidl, H.R.; Stollé, C.A.; Seyer, J.M.

J. Biol. Chem. 260, 5533-5541, 1985

A;Title: Partial covalent structure of the human alpha 2 type V collagen chain.

A;Reference number: I55239; MUID:95182703; PMID:2985598

A;Accession: I55239

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1002-1226 <RE2>

A;Cross-references: UNIPARC:UPI0000072247; GB:M10956; NID:G180427; PIDN:AAA52007.1; PID:

A;Note: part of this sequence were determined by protein sequencing

R;Emanuel, B.S.; Cannizzaro, L.A.; Seyer, J.M.; Myers, J.C.

Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985

A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm

A;Reference number: I59025; MUID:85216505; PMID:3858826

A;Accession: I59025

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1003-1034 <RES>

A;Cross-references: UNIPARC:UPI000016A627; GB:M11135; NID:G179693; PIDN:AAA51857.1; PID:

A;Note: part of this sequence were determined by protein sequencing

R;Myers, J.C.; Loidl, H.R.; Seyer, J.M.; Dion, A.S.

J. Biol. Chem. 260, 11216-11222, 1985

A;Title: Complete primary structure of the human alpha-2 type V procollagen COOH-termina

A;Reference number: A25374; MUID:85289337; PMID:2411731

A;Accession: A25374

A;Molecule type: mRNA

A;Residues: 1227-1417, 'T', 1419-1437, 'S', 1439-1496 <MYE>

A;Cross-references: UNIPARC:UPI000016A721; GB:M11718; NID:G180912; PIDN:AAA52058.1; PID:

A;Experimental source: normal fibroblasts

R;Tsipouras, P.; Schwartz, R.C.; Liddell, A.C.; Salkeld, C.S.; Weil, D.; Ramirez, F.

Genomics 3, 275-277, 1988

A;Title: Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2, located on

A;Reference number: A30017; MUID:89138450; PMID:3224983

A;Accession: A30017
A;Molecule type: DNA
A;Residues: 1449-1463, 'E', 1465-1495, 'A' <TSI>
A;Cross-references: UNIPARC:UPI0000173B95; GB:J03051; NID:G179695; PIDN:AAAS1858.1; PID:
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Note: the authors translated the codon GAA for residue 1460 as Gln, and GAG for residue
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit
are 5-hydroxylated and subsequently O-glycosylated.
C;Comment: The amino-terminal propeptide domain appears not to be completely cleaved.
C;Genetics:
A;Gene: GDB:COL5A2
A;Cross-references: GDB:119064; OMIM:120190
A;Map position: 2q31-2q31
A;Introns: 33/1; 812/3; 830/3; 848/3; 902/3; 922/3; 974/3; 1046/3; 1064/3; 1448/3
C;Complex: type V collagen may be a homotrimer of alpha 1(IV) chains (see PIR:CGHUIV), a
alpha 2(IV) chain and one alpha 3(V) chain, initially linked by disulfide bonds among the
length, is formed with desmosine cross-links made from lysine and allysine residues
C;Function:
A;Description: structural component of extracellular fibrous polymer associated with cell
A;Note: may play a role in controlling the lateral growth of collagen I fibrils
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F;1-26/Domain: signal sequence #status predicted <SIG>
F;127-1250/Product: collagen alpha 2(V) chain #status predicted <MAT>
F;127-193/Domain: amino-terminal propeptide (uncleaved) #status predicted <NPP>
F;127-108/Region: nonhelical
F;140-99/Domain: von Willebrand factor type C repeat homology <VWC>
F;109-186/Region: helical
F;187-208/Region: nonhelical
F;209-1225/Region: helical
F;209-1225/Region: cell attachment (R-G-D) motif
F;941-943/Region: cell attachment (R-G-D) motif
F;1064-1066/Region: cell attachment (R-G-D) motif
F;1067-1069/Region: cell attachment (R-G-D) motif
F;1097-1099/Region: cell attachment (R-G-D) motif
F;1124-1126/Region: cell attachment (R-G-D) motif
F;1133-1135/Region: cell attachment (R-G-D) motif
F;1225-1250/Region: carboxyl-terminal nonhelical telopeptide
F;1251-1496/Domain: carboxyl-terminal propeptide #status predicted <CPP>
F;1269-1496/Domain: fibrillar collagen propeptide #status predicted <FCC>
F;127/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;193-194/Cleavage site: Ala-Gln (procollagen N-endopeptidase) #status predicted
F;194/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
F;201/Modified site: allysine (Lys) #status predicted
F;290,293,296,608,614,1004,1007,1013,1028,1034/Modified site: 4-hydroxyproline (Pro) #st
F;299,1139/Modified site: 5-hydroxylysine (Lys) #status predicted
F;299,1139/Binding site: carboxylate (Lys) (covalent) #status predicted
F;1025/Modified site: 5-hydroxylysine (Lys) #status experimental
F;1250-1251/Cleavage site: Glu-Asp (procollagen C-endopeptidase)
F;1259,1397/Binding site: carboxylate (Asn) (covalent) #status predicted
F;1293,1299,1325/Disulfide bonds: interchain #status predicted
F;1333-1494,1402-1447/Disulfide bonds: #status predicted

Query Match 69.4%; Score 43; DB 1; Length 1496;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGGSPG 10
Db 780 GEKGAGGTAG 789

RESULT 70
149607
A;Accession: 149607
C;Superfamily: collagen type V alpha 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: 149607
R;Andrikopoulos, K.; Suzuki, H.R.; Solursh, M.; Ramirez, F.
Dev. Dyn. 195, 113-120, 1992
A;Title: Localization of pro-alpha 2(V) collagen transcripts in the tissues of the devel
A;Reference number: 149607; MUID:93214071; PMID:1297453
A;Accession: 149607
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA
A;Residues: 1-1497 <RES>
A;Cross-references: UNIPROT:Q61431; UNIPARC:UPI0000028657; GB:L02918; NID:G309180; PIDN:
C;Genetics:
A;Gene: COL5A-2
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;33-98/Domain: von Willebrand factor type C repeat homology <VWC>
F;1270-1497/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
Query Match 69.4%; Score 43; DB 2; Length 1497;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGGSPG 10
Db 781 GEKGAGGTAG 790

RESULT 71
CGHUEB
A;Accession: A54122
A;Molecule type: mRNA
A;Residues: 1-1691 <ZHO>
A;Cross-references: UNIPROT:Q14031; UNIPARC:UPI0000161C20; GB:U04845; NID:G496977; PIDN:
R;Ohashi, T.; Sugimoto, M.; Mattei, M.G.; Ninomiya, Y.
J. Biol. Chem. 269, 7520-7526, 1994
A;Title: Identification of a new collagen IV chain, alpha6(IV), by cDNA isolation and as
A;Reference number: A53404; MUID:94171779; PMID:8125972
A;Accession: A53404
A;Molecule type: mRNA
A;Residues: 'MHPG', 6-169, 'M', 171-916, 'S', 918-1301, 1314-1355, 'A', 1357-1691 <OOH>
A;Cross-references: UNIPARC:UPI000016A6F1; DBJ:D21337; NID:G466537; PIDN:BA04809.1; PI
R;Zhou, J.; Mochizuki, T.; Smets, H.; Antignac, C.; Laurila, P.; de Paape, A.; Tryggvas
Science 261, 1167-1169, 1993
A;Title: Deletion of the paired alphas(IV) and alpha6(IV) collagen genes in inherited sr
A;Reference number: A57079; MUID:93361972; PMID:8356449
A;Accession: B57079
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-546, 'G' <ZH2>
A;Cross-references: UNIPARC:UPI0000147AB2; GB:L22763
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
ed and subsequently O-glycosylated.
C;Genetics:
A;Gene: GDB:COL4A6
A;Cross-references: GDB:222775; OMIM:303631
A;Map position: Xq22-Xq22
A;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5(
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-l
C;Function:
A;Description: minor structural component of extracellular basement membrane
C;Superfamily: collagen alpha 1(IV) chain
C;Keywords: Alport syndrome; basement membrane; coiled coil; extracellular matrix; glyco
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1691/Product: collagen alpha 6(IV) chain #status predicted <MAT>
F;22-46/Domain: amino-terminal nonhelical, NC2 <NC2>
F;47-1463/Region: interrupted helical
F;1464-1691/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F;1473-1571/Domain: collagen IV carboxyl-terminal repeat <CT1>
F;1581-1687/Domain: collagen IV carboxyl-terminal repeat <CT2>
F;31,36,40,42,126,482,484,657/Disulfide bonds: interchain #status predicted
F;127/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:1482-1568,1515-1571/Disulfide bonds: (or 1482-1571, 1515-1568) #status predicted
 F:1527-1533,1636-1643/Disulfide bonds: #status predicted
 F:1590-1684,1624-1687/Disulfide bonds: (or 1590-1687, 1624-1684) #status predicted

Query Match 69.4%; Score 43; DB 1; Length 1691;
 Best Local Similarity 72.7%; Pred. No. 42;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
 ||||| : |||||
 Db 288 GEKGAGSPGL 298

RESULT 72

collagen alpha-4 chain precursor - sea urchin (Strongylocentrotus purpuratus)
 N;Alternate names: collagen alpha 2(IV) chain homolog
 C;Species: Strongylocentrotus purpuratus (purple urchin)
 C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004
 C;Accession: A54121; S44317
 R;Exposito, J.Y.; Suzuki, H.; Geourjon, C.; Garrone, R.; Solursh, M.; Ramirez, F.
 J. Biol. Chem. 269, 13167-13171, 1994
 A;Title: Identification of a cell lineage-specific gene coding for a sea urchin alpha2(I)
 A;Reference number: A54121; MUID:94230414; PMID:8175744
 A;Accession: A54121
 A;Molecule type: mRNA
 A;Residues: 1-1747 <EXP>
 A;Cross-references: UNIPROT:Q26640; UNIPARC:UPI0000076901; EMBL:X76730; NID:9483606; PID
 C;Genetics:
 A;Gene: COLP4alpha
 C;Superfamily: collagen alpha 1(IV) chain

Query Match 69.4%; Score 43; DB 2; Length 1747;
 Best Local Similarity 63.6%; Pred. No. 44;
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
 ||||| : |||||
 Db 1395 GEKGAGSPGL 1405

RESULT 73

collagen alpha 2(IV) chain precursor - pig roundworm
 C;Species: Ascaris suum (pig roundworm)
 C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 09-Jul-2004
 C;Accession: S16366
 R;Pettit, J.; Kingston, I.B.
 J. Biol. Chem. 266, 16149-16156, 1991
 A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti
 A;Reference number: S16366; MUID:91340768; PMID:1714907
 A;Accession: S16366
 A;Molecule type: mRNA
 A;Residues: 1-1763 <BJ>
 A;Cross-references: UNIPROT:P27393; UNIPARC:UPI0000126D40; GB:M67507; NID:9159648; PIDN:
 C;Genetics:
 A;Intons: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/1;
 C;Superfamily: collagen alpha 1(IV) chain
 C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disulfid
 F;1-26/Domain: signal sequence #status predicted <SIG>
 F;27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
 F;27-42/Domain: non-collagenous NH1 #status predicted <NH1>
 F;43-1529/Domain: collagenous #status predicted <COI>
 F;197-199/Region: cell attachment (R-G-D) motif
 F;1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
 F;1530-1638/Domain: repeat NC1 #status predicted <NC11>
 F;1639-1763/Domain: repeat NC1 #status predicted <NC12>
 F;31.34.39.41.536.539/Disulfide bonds: interchain #status predicted
 F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 69.4%; Score 43; DB 2; Length 1763;
 Best Local Similarity 72.7%; Pred. No. 44;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPGL 11
 ||||| : |||||
 Db 784 GEKGAGSPGL 794

RESULT 74

A54849
 collagen alpha 1(VII) chain precursor - human
 N;Alternate names: procollagen alpha 1(VII) chain
 C;Species: Homo sapiens (man)
 C;Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 09-Jul-2004
 C;Accession: A54849; PH0844; S16316; I56328; A30296; I84686
 R;Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
 J. Biol. Chem. 269, 20256-20262, 1994
 A;Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII)
 A;Reference number: A54849; MUID:94327588; PMID:8051117
 A;Accession: A54849
 A;Status: not compared with conceptual translation
 A;Molecule type: mRNA
 A;Residues: 1-2944 <CHR>
 A;Cross-references: UNIPROT:Q02389; UNIPARC:UPI000017A138; GB:L02870; NID:9987124; PIDN:
 R;Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
 Biochem. Biophys. Res. Commun. 183, 958-963, 1992
 A;Title: Molecular cloning and characterization of type VII collagen cDNA.
 A;Reference number: PH0844; MUID:92231902; PMID:1567409
 A;Accession: PH0844
 A;Molecule type: mRNA
 A;Residues: 'EPR', 340-475, 'RALSTASHSTLCWRATRHPCNRGSHWTRACPCNPASHRAARAG', 524-528, 'C',
 A;Cross-references: UNIPARC:UPI000017A139; DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BA
 A;Experimental source: Keratinocyte
 A;Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
 R;Parente, M.G.; Chung, L.S.A.; Rynnaenen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat
 Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
 A;Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
 A;Reference number: S16316; MUID:91334380; PMID:1871109
 A;Accession: S16316
 A;Molecule type: mRNA
 A;Residues: 815-892, 'E', 894-1439 <PAR>
 A;Cross-references: UNIPARC:UPI000016A722; GB:M65158; GB:S49017; NID:9180914; PIDN:AAA96
 A;Experimental source: Keratinocyte
 R;Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prissayanh, P.S.; Cook, M.E.; Wright, J.;
 J. Invest. Dermatol. 99, 691-696, 1992
 A;Title: Noncollagenous (NCI) domain of collagen VII resembles multidomain adhesion prote
 A;Reference number: I56328; MUID:93107742; PMID:1469284
 A;Accession: I56328
 A;Status: translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>
 A;Cross-references: UNIPARC:UPI000016B3AC; GB:S51236; NID:9262308; PIDN:AAB24637.1; PID:
 R;Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.
 J. Biol. Chem. 264, 3822-3826, 1989
 A;Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagena
 A;Reference number: A30296; MUID:89139437; PMID:2537292
 A;Accession: A30296
 A;Molecule type: protein
 A;Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041;
 A;Cross-references: UNIPARC:UPI000017A13A; UNIPARC:UPI000017A13B; UNIPARC:UPI000017A13C;
 A;Note: two reported peptides cannot be reliably located
 R;Greenspan, D.S.
 Hum. Mol. Genet. 2, 273-278, 1993
 A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous
 A;Reference number: 148103; MUID:93271985; PMID:8499916
 A;Accession: I84686
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 2395-2871, 'S', 2873-2944 <RE2>
 A;Cross-references: UNIPARC:UPI000016A724; GB:L06862; NID:9388713; PIDN:AAA89196.1; PID:
 R;Christiano, A.M.; Rynnaenen, M.; Uitto, J.
 Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
 A;Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser subs
 A;Reference number: A55255; MUID:94224777; PMID:8170945

A:Contents: annotation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL7A1; EBR1; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermolys
A:Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
C:Function:
A:Description: structural component of extracellular polymer associated with anchoring f
A:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproli
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NCL>
F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>
F:231-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-593/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-2783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>
F:337,786,1109/Binding site: carboxylate (Asn) (covalent) #status predicted
F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2625,2631/Binding site: carboxylate (Lys) (covalent) #status experimental
F:2634,2802,2804/Disulfide bonds: interchain #status predicted

Query Match 69.4%; Score 43; DB 2; Length 2944;
Best Local Similarity 80.0%; Pred. No. 73;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSGPG 10
|||||
Db 1637 GEKGDEGPPG 1646

RESULT 75
S18250
collagen alpha 1(II) chain precursor - chicken (fragment)
C:Species: Gallus gallus (chicken)
C:Date: 22-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 09-Jul-2004
C:Accession: S18250
R:Nah, H.D.; Upholt, W.B.
J. Biol. Chem. 266, 23446-23452, 1991
A:Title: Type II collagen mRNA containing an alternatively spliced exon predominates in
A:Reference number: S18250; MUID:92078225; PMID:1744138
A:Accession: S18250
A:Molecule type: mRNA
A:Residues: 1-206 <NAH>
A:Cross-references: UNIPROT:Q90802; UNIPARC:UPI00000PDSAA; EMBL:M74435; NID:9211635; PID
C:Genetics:
A:Gene: COL2A1
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C:Keywords: alternative splicing; coiled coil; extracellular matrix; glycoprotein
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-184/Domain: amino-terminal propeptide #status predicted <PRO>
F:36-95/Domain: von Willebrand factor type C repeat homology <VWC>
F:185-206/Product: collagen alpha 1(II) chain (fragment) #status predicted <MAT>

Query Match 67.7%; Score 42; DB 2; Length 206;
Best Local Similarity 70.0%; Pred. No. 7.9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSGPG 10
|||||
Db 141 GEKGEGKAPG 150

Search completed: March 11, 2006, 12:06:23
Job time : 32.2 secs

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 11:52:06 ; Search time 175.8 Seconds
(without alignments)
48.159 Million cell updates/sec

Title: US-10-698-121A-2
Perfect score: 62
Sequence: 1 GKGAGSGPGLL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58	93.5	584	2	Q9228 HUMAN
2	58	93.5	623	2	Q14035 HUMAN
3	58	93.5	652	2	Q5AT5 HUMAN
4	58	93.5	683	2	Q5AT5 HUMAN
5	58	93.5	717	2	Q5AT5 HUMAN
6	58	93.5	717	2	Q5AT5 HUMAN
7	58	93.5	717	2	Q5AT5 HUMAN
8	56	90.3	568	2	Q8CD80 MOUSE
9	56	90.3	739	2	Q70575 MOUSE
10	56	90.3	751	2	Q9RIN9 MOUSE
11	52	83.9	447	2	Q16593 HUMAN
12	52	83.9	1463	2	Q59F89 HUMAN
13	52	83.9	1603	1	COGAL HUMAN
14	51	82.3	1046	2	Q8K388 MOUSE
15	51	82.3	1580	2	Q8BLX7 MOUSE
16	50	80.6	396	1	CO9A2 HUMAN
17	49	79.0	396	2	Q619U9 CAENORHABDI
18	49	79.0	518	2	Q5C3P1 SCHJEA
19	49	79.0	650	2	Q17866 CAENORHABDI
20	49	79.0	662	2	Q4RGB0 TETRAODON
21	49	79.0	680	1	COAAL MOUSE
22	49	79.0	778	2	Q9U9K6 CAENORHABDI
23	49	79.0	1117	2	Q9U9K7 CAENORHABDI
24	49	79.0	1138	2	Q7JL30 CAENORHABDI
25	49	79.0	1175	2	Q810G6 CAENORHABDI
26	49	79.0	1706	2	Q5VZA9 HUMAN
27	49	79.0	1712	1	COA42 HUMAN
28	49	79.0	1736	1	COBA2 MOUSE
29	49	79.0	1748	2	Q591P1 PIG
30	49	79.0	2936	2	Q7YRK8 CANFA
31	48	77.4	60	2	Q25617 ONCHOCERCA

32	48	77.4	132	2	P78429 HUMAN
33	48	77.4	338	2	Q8BJU6 MOUSE
34	48	77.4	636	1	CO3A1 RAT
35	48	77.4	660	2	Q86D04 CAEBL
36	48	77.4	1049	1	CO3A1 BOVIN
37	48	77.4	1163	2	Q8N6U4 HUMAN
38	48	77.4	1222	2	Q8K173 MOUSE
39	48	77.4	1258	2	Q8AW11 BRAKE
40	48	77.4	1463	2	Q5PQT6 RAT
41	48	77.4	1464	1	CO3A1 MOUSE
42	48	77.4	1464	2	Q8BKV2 MOUSE
43	48	77.4	1464	2	Q8BLW4 MOUSE
44	48	77.4	1464	2	Q7TT32 MOUSE
45	48	77.4	1466	1	CO3A1 HUMAN
46	48	77.4	1466	2	Q53S91 HUMAN
47	48	77.4	1466	2	Q541P8 HUMAN
48	48	77.4	1467	2	Q5DTG2 MOUSE
49	47	75.8	243	2	Q95JD7 MACMU
50	47	75.8	244	1	ADIPG HUMAN
51	47	75.8	244	2	Q58EX9 HUMAN
52	47	75.8	375	2	Q5TMG9 ANOGEA
53	47	75.8	802	2	Q7PYX1 ANOGEA
54	47	75.8	856	2	Q4SB07 TETNG
55	47	75.8	1026	2	Q4S273 TETNG
56	47	75.8	1027	1	CAFF RIFPA
57	47	75.8	1136	2	Q35053 MOUSE
58	47	75.8	1333	2	Q4SK58 TETNG
59	47	75.8	1549	2	Q60444 CRIGR
60	47	75.8	1685	1	CO4A5 HUMAN
61	47	75.8	1685	2	Q9NUB7 HUMAN
62	46	74.2	53	2	Q16299 HUMAN
63	46	74.2	142	2	Q866A5 PIG
64	46	74.2	167	2	Q9TT85 PIG
65	46	74.2	184	2	Q92416 CAVPO
66	46	74.2	187	2	Q9GLX9 RABIT
67	46	74.2	194	2	Q95J95 CANFA
68	46	74.2	198	2	Q5JPA6 HUMAN
69	46	74.2	229	2	Q4Z8Q0 ALOLA
70	46	74.2	229	2	Q4Z8Q1 NYCPR
71	46	74.2	310	2	Q90612 CHICK
72	46	74.2	347	2	Q9XT25 BOVIN
73	46	74.2	441	1	EMID2 HUMAN
74	46	74.2	546	2	Q6P7U1 MOUSE
75	46	74.2	546	2	Q99K97 MOUSE
76	46	74.2	771	2	Q61DU2 CAEBR
77	46	74.2	799	2	Q8BNS7 MOUSE
78	46	74.2	812	2	Q06452 EPHMU
79	46	74.2	826	2	Q8K0N6 MOUSE
80	46	74.2	886	2	Q8CEF7 MOUSE
81	46	74.2	888	2	Q90796 CHICK
82	46	74.2	911	1	COBAl BOVIN
83	46	74.2	920	2	Q78EC6 SMURI
84	46	74.2	977	2	Q4SEP8 TETNG
85	46	74.2	1009	2	Q5TUG3 CANFA
86	46	74.2	1053	2	Q92084 MOUSE
87	46	74.2	1208	2	Q4RX03 TETNG
88	46	74.2	1262	1	CO3A1 CHICK
89	46	74.2	1269	2	Q7T2Z7 CHICK
90	46	74.2	1347	2	Q4S1F4 TETNG
91	46	74.2	1418	1	CO2A1 HUMAN
92	46	74.2	1418	2	Q28396 HORSE
93	46	74.2	1419	2	Q80VY3 MOUSE
94	46	74.2	1419	2	Q80X38 MOUSE
95	46	74.2	1419	2	Q63123 RAT
96	46	74.2	1420	2	Q90W37 CHICK
97	46	74.2	1442	2	Q62031 MOUSE
98	46	74.2	1442	2	Q62033 MOUSE
99	46	74.2	1459	1	CO2A1 MOUSE
100	46	74.2	1459	2	Q62032 MOUSE
101	46	74.2	1475	2	Q4S5M8 TETNG
102	46	74.2	1487	2	Q14047 HUMAN
103	46	74.2	1487	2	Q77753 CANFA
104	46	74.2	1487	2	Q641K3 MOUSE

P78429	homo sapien
Q8bjue	mus musculus
P13941	rattus norv
Q86d04	caenorhabdi
P04258	bos taurus
Q8n6u4	homo sapien
Q8k173	mus musculus
Q8aw11	brachydanio
Q5pq66	rattus norv
P08121	mus musculus
Q8blw4	mus musculus
Q7tt32	mus musculus
P02461	homo sapien
Q53s91	homo sapien
Q541p8	homo sapien
Q5dtg2	mus musculus
Q95jd7	macaca mula
Q15848	homo sapien
Q58ex9	homo sapien
Q5tmg9	anopheles g
Q7pyx1	anopheles g
Q4sb07	tetradon n
Q4s273	tetradon n
P30754	riftia pach
Q35053	mus musculus
Q4sk58	tetradon n
Q60444	cricetus
P29400	homo sapien
Q9nub7	homo sapien
Q16299	homo sapien
Q866a5	sus scrofa
Q9tt85	sus scrofa
Q92416	cavia porce
Q9glx9	oryctolagus
Q95j95	canis famil
Q5jpa6	homo sapien
Q4z8q0	alopex lego
Q4z8q1	nyctereutes
Q90612	gallus gall
Q9xt25	bos taurus
Q6a83	homo sapien
Q6p7u1	mus musculus
Q99k97	mus musculus
Q61du2	caenorhabdi
Q8bns7	mus musculus
Q06452	ephydatia m
Q8k0n6	mus musculus
Q8cef7	mus musculus
Q90796	gallus gall
Q28083	bos taurus
Q78ec6	mus sp. typ
Q4sep8	tetradon n
Q5cj93	canis famil
Q92084	mus musculus
Q4rx03	tetradon n
P12105	gallus gall
Q7t2z7	gallus gall
Q4slp4	tetradon n
P02458	homo sapien
Q28396	equus caball
Q80vy3	mus musculus
Q80x38	mus musculus
Q63123	rattus norv
Q90w37	gallus gall
Q62031	mus musculus
Q62033	mus musculus
P28481	mus musculus
Q62032	mus musculus
Q62032	mus musculus
Q4s5m8	tetradon n
Q14047	homo sapien
Q77753	canis famil
Q641k3	mus musculus

105	46	74.2	1491	2	Q91718_XENLA	Q91718 xenopus lae	178	44	71.0	396	2	Q22651_CAEBL	Q22651 caenorhabdi
106	45	74.2	1491	2	Q7ZTM4_XENLA	Q7ZTM4 xenopus lae	179	44	71.0	568	2	Q6NUV1_BRARE	Q6NUV1 brachydanio
107	46	74.2	1492	2	Q6P422_XENTR	Q6P422 xenopus tro	180	44	71.0	589	2	Q91LL6_MOUSE	Q91LL6 mus musculus
108	46	74.2	1562	2	Q6GQS7_MOUSE	Q6GQS7 mus musculus	181	44	71.0	673	2	Q14052_HUMAN	Q14052 homo sapien
109	46	74.2	1596	2	Q5TJG0_CANFA	Q5TJG0 canis fami	182	44	71.0	675	2	Q9N178_PIG	Q9N178 sus scrofa
110	46	74.2	1650	2	Q5JP94_HUMAN	Q5JP94 homo sapien	183	44	71.0	680	1	COA11_HUMAN	Q03692 homo sapien
111	46	74.2	1658	2	Q59GD4_HUMAN	Q59GD4 homo sapien	184	44	71.0	681	1	Q6AQZ8_XENLA	Q6AQZ8 xenopus lae
112	46	74.2	1669	1	CO4A1_MOUSE	Q2463 mus musculus	185	44	71.0	688	1	CO9A2_MOUSE	Q07643 mus musculus
113	46	74.2	1684	2	Q8HYC1_CANFA	Q8HYC1 canis fami	186	44	71.0	688	2	Q8K2W0_MOUSE	Q8K2W0 mus musculus
114	46	74.2	1691	1	CO4A5_CANFA	Q28247 canis fami	187	44	71.0	698	2	Q867Z8_BACRY	Q867Z8 bactrocera
115	46	74.2	1691	2	Q9ESQ1_MOUSE	Q9ESQ1 mus musculus	188	44	71.0	747	1	CO2A1_BOVIN	Q02459 bos taurus
116	46	74.2	1691	2	Q9ESQ2_MOUSE	Q9ESQ2 mus musculus	189	44	71.0	763	2	Q4RZL2_TETNG	Q4RZL2 tetraodon n
117	46	74.2	1691	2	Q63ZW6_MOUSE	Q63ZW6 mus musculus	190	44	71.0	905	2	Q8MW55_MYTGA	Q8MW55 mytilus gal
118	46	74.2	1693	2	Q5STP6_HUMAN	Q5STP6 homo sapien	191	44	71.0	998	2	Q8CFM4_MOUSE	Q8CFM4 mus musculus
119	46	74.2	1693	2	Q5SU18_HUMAN	Q5SU18 homo sapien	192	44	71.0	1014	2	Q5FWW2_XENLA	Q5FWW2 xenopus lae
120	46	74.2	1736	1	COBA2_HUMAN	Q1942 homo sapien	193	44	71.0	1372	1	CO1A2_RAT	Q02466 rattus norv
121	46	74.2	1737	2	Q9J104_RAT	Q9J104 mus musculus	194	44	71.0	1418	2	Q9W7R9_CYNPY	Q9W7R9 cynops pyrr
122	46	74.2	1739	2	Q9JL12_MOUSE	Q9JL12 mus musculus	195	44	71.0	1438	2	Q4S012_TETNG	Q4S012 tetraodon n
123	46	74.2	1745	1	CO5A3_HUMAN	Q25940 homo sapien	196	44	71.0	1464	2	Q6P912_XENLA	Q6P912 xenopus lae
124	46	74.2	1759	1	COBA1_CAEBL	P17139 caenorhabdi	197	44	71.0	1472	2	Q90ZA0_CHICK	Q90ZA0 gallus gall
125	46	74.2	1804	1	COBA1_MOUSE	Q61245 mus musculus	198	44	71.0	1617	2	Q6MGB2_RAT	Q6MGB2 rattus norv
126	46	74.2	1804	2	Q80WR4_MOUSE	Q80WR4 mus musculus	199	44	71.0	1669	2	Q9QZS0_MOUSE	Q9QZS0 mus musculus
127	46	74.2	1806	1	COBA1_HUMAN	Q12107 homo sapien	200	44	71.0	1688	2	Q597P9_CANFA	Q597P9 canis fami
128	46	74.2	1806	2	Q5VT31_HUMAN	Q5VT31 homo sapien	201	44	71.0	1707	1	CO4A2_MOUSE	P08122 mus musculus
129	46	74.2	2944	2	Q63870_MOUSE	Q63870 mus musculus	202	44	71.0	1816	2	Q6NTI5_BRARE	Q6NTI5 brachydanio
130	45	72.6	289	2	Q60QP0_CAEBR	Q60QP0 caenorhabdi	203	43	69.4	62	2	Q6RV17_BOSMU	Q6RV17 bos mutus g
131	45	72.6	289	2	Q61YU7_CAEBR	Q61YU7 caenorhabdi	204	43	69.4	82	2	Q77539_CANFA	Q77539 canis fami
132	45	72.6	291	2	Q61YU7_CAEBR	Q61YU7 caenorhabdi	205	43	69.4	115	2	Q61428_MOUSE	Q61428 mus musculus
133	45	72.6	291	2	Q9NAR3_CAEBL	Q9NAR3 caenorhabdi	206	43	69.4	133	2	Q7KZ49_HUMAN	Q7KZ49 homo sapien
134	45	72.6	301	2	Q7SZ01_XENLA	Q7SZ01 xenopus lae	207	43	69.4	140	2	Q13995_HUMAN	Q13995 homo sapien
135	45	72.6	323	2	Q71RP4_HUMAN	Q71RP4 homo sapien	208	43	69.4	141	2	Q13993_HUMAN	Q13993 homo sapien
136	45	72.6	440	1	EMID2_MOUSE	Q71RF4 mus musculus	209	43	69.4	152	2	Q13996_HUMAN	Q13996 homo sapien
137	45	72.6	473	2	Q4SB10_TETNG	Q4SB10 tetraodon n	210	43	69.4	152	2	Q21855_CAEBL	Q21855 caenorhabdi
138	45	72.6	486	2	Q4T415_TETNG	Q4T415 tetraodon n	211	43	69.4	153	2	Q13994_HUMAN	Q13994 homo sapien
139	45	72.6	723	1	VND_DROME	P22808 drosophila	212	43	69.4	167	2	Q13992_HUMAN	Q13992 homo sapien
140	45	72.6	867	2	Q59EX5_HUMAN	Q59EX5 mus sapien	213	43	69.4	183	2	Q62723_PIG	Q62723 sus scrofa
141	45	72.6	894	2	Q8MW54_MYTGA	Q8MW54 mytilus gal	214	43	69.4	201	2	Q9NT93_HUMAN	Q9NT93 homo sapien
142	45	72.6	904	2	Q76271_MYTED	Q76271 mytilus edu	215	43	69.4	242	2	Q5GAJ6_9VIRU	Q5GAJ6 grouper iri
143	45	72.6	907	2	Q26312_STRPU	Q26312 strongyloce	216	43	69.4	245	2	Q7KZ33_HUMAN	Q7KZ33 homo sapien
144	45	72.6	954	2	Q8WXV8_HUMAN	Q8WXV8 homo sapien	217	43	69.4	255	2	Q4SDS9_TETNG	Q4SDS9 tetraodon n
145	45	72.6	957	2	Q96P44_HUMAN	Q96P44 homo sapien	218	43	69.4	271	2	Q5U3G1_BRARE	Q5U3G1 brachydanio
146	45	72.6	957	2	Q9H0V3_HUMAN	Q9H0V3 homo sapien	219	43	69.4	271	2	Q7T0T0_XENLA	Q7T0T0 xenopus lae
147	45	72.6	1009	2	Q4SNW2_TETNG	Q4SNW2 tetraodon n	220	43	69.4	280	2	Q91014_CHICK	Q91014 gallus gall
148	45	72.6	1069	2	Q6LAN8_HUMAN	Q6LAN8 homo sapien	221	43	69.4	281	2	Q616H5_CAEBR	Q616H5 caenorhabdi
149	45	72.6	1075	2	Q86X41_HUMAN	Q86X41 homo sapien	222	43	69.4	281	2	Q629B6_CAEBR	Q629B6 caenorhabdi
150	45	72.6	1115	2	Q6PFB1_MOUSE	Q6PFB1 mus musculus	223	43	69.4	281	2	Q44174_CAEBL	Q44174 caenorhabdi
151	45	72.6	1225	2	Q6PCL3_MOUSE	Q6PCL3 mus musculus	224	43	69.4	281	2	Q44774_CAEBL	Q44774 caenorhabdi
152	45	72.6	1414	2	Q26634_STRPU	Q26634 strongyloce	225	43	69.4	290	2	Q26054_PARLI	Q26054 paracentrot
153	45	72.6	1453	1	CO1A1_MOUSE	P11087 mus musculus	226	43	69.4	294	2	Q19979_CAEBL	Q19979 caenorhabdi
154	45	72.6	1453	2	Q63079_RAT	Q63079 rattus norv	227	43	69.4	317	2	Q4W514_HUMAN	Q4W514 homo sapien
155	45	72.6	1453	2	Q810J9_MOUSE	Q810J9 mus musculus	228	43	69.4	320	2	Q61416_CAEBR	Q61416 caenorhabdi
156	45	72.6	1460	1	CO1A1_CANFA	Q9XSJ7 canis fami	229	43	69.4	323	2	Q7M099_MESAU	Q7M099 mesoicetu
157	45	72.6	1461	2	Q76045_HUMAN	Q76045 homo sapien	230	43	69.4	327	2	Q27276_CAEBL	Q27276 caenorhabdi
158	45	72.6	1464	1	CO1A1_HUMAN	P02452 homo sapien	231	43	69.4	335	1	FCN1_RAT	Q9WT88 rattus norv
159	45	72.6	1464	2	Q8N473_HUMAN	Q8N473 homo sapien	232	43	69.4	335	2	Q5M8B4_RAT	Q5M8B4 rattus norv
160	45	72.6	1467	2	Q59FF4_HUMAN	Q59FF4 homo sapien	233	43	69.4	340	2	Q6DJ36_XENTR	Q6DJ36 xenopus tro
161	45	72.6	1655	2	Q4SDF0_TETNG	Q4SDF0 tetraodon n	234	43	69.4	358	2	Q23222_CAEBL	Q23222 caenorhabdi
162	45	72.6	1655	1	CO4A1_HUMAN	P02462 homo sapien	235	43	69.4	393	2	Q60ZN4_CAEBR	Q60ZN4 caenorhabdi
163	45	72.6	1669	2	Q5VWF6_HUMAN	Q5VWF6 mus musculus	236	43	69.4	405	2	Q9N3X9_CAEBL	Q9N3X9 caenorhabdi
164	45	72.6	1682	2	Q9QZ39_MOUSE	Q9QZ39 mus musculus	237	43	69.4	428	2	Q4RZT4_TETNG	Q4RZT4 tetraodon n
165	45	72.6	1714	2	Q7Z515_HUMAN	Q7Z515 homo sapien	238	43	69.4	418	2	Q5FY72_CHICK	Q5FY72 gallus gall
166	45	72.6	1714	2	QSVY50_HUMAN	QSVY50 homo sapien	239	43	69.4	419	1	COLE_LEPMA	Q98085 lepomis mac
167	45	72.6	1723	2	Q9GQB1_HYDAT	Q9GQB1 hydra atten	240	43	69.4	438	2	Q7M3U7_9ANNE	Q7M3U7 paravivell
168	45	72.6	1752	2	Q07265_STRPU	Q07265 strongyloce	241	43	69.4	444	2	Q4SEW7_TETNG	Q4SEW7 tetraodon n
169	45	72.6	1775	1	CO4A1_DROME	P08120 drosophila	242	43	69.4	458	2	Q9NAB3_CAEBL	Q9NAB3 caenorhabdi
170	44	71.0	95	2	Q5NC68_MOUSE	Q5NC68 mus musculus	243	43	69.4	469	2	Q70598_RAT	Q70598 rattus norv
171	44	71.0	164	2	Q5BW21_SCHJA	Q5BW21 schistosoma	244	43	69.4	473	2	Q70605_RAT	Q70605 rattus norv
172	44	71.0	198	2	Q9DDR4_9SAUR	Q9DDR4 chrysemya s	245	43	69.4	482	1	COBA1_RAT	COBA1 RAT
173	44	71.0	217	1	PGRC2_RAT	Q9DDF4 chrysemya s	246	43	69.4	497	2	Q51H10_MAGGR	Q51H10 magnaporth
174	44	71.0	229	2	Q9VE77_DROME	Q9VE77 drosophila	247	43	69.4	515	2	Q4SIU3_TETNG	Q4SIU3 tetraodon n
175	44	71.0	230	2	Q9R149_CAVPO	Q9R149 cavia porce	248	43	69.4	520	1	MARCO_HUMAN	Q9UEW3 homo sapien
176	44	71.0	266	2	Q4SIU5_TETNG	Q4SIU5 tetraodon n	249	43	69.4	520	2	Q4ZG40_HUMAN	Q4ZG40 homo sapien
177	44	71.0	292	2	Q9CSB7_MOUSE	Q9CSB7 mus musculus	250	43	69.4	532	2	Q810Y4_RAT	Q810Y4 rattus norv

251	43	69.4	532	2	Q8K4G2_MOUSE	Q8K4G2_mus musculus	324	42	67.7	165	2	Q73PN7_TREDE	Q73pn7 treponema d
252	43	69.4	532	2	Q5SUQ0_MOUSE	Q5sug0_mus musculus	325	42	67.7	184	2	Q5ITE6_ANOGA	Q5tte6 anopheles g
253	43	69.4	532	2	Q86Y22_HUMAN	Q86y22_homo sapien	326	42	67.7	187	2	Q9BDZ2_MACMU	Q9bdz2 macaca mula
254	43	69.4	579	2	Q4RQ87_TETNG	Q4rq87_tetradodon n	327	42	67.7	188	2	Q7PH20_ANOGA	Q7ph20 anopheles g
255	43	69.4	600	2	Q824A0_CAEBR	Q824a0_caenorhabdi	328	42	67.7	206	2	Q90802_CHICK	Q90802 gallus gall
256	43	69.4	610	2	Q4RML4_TETNG	Q4rml4_tetradodon n	329	42	67.7	210	2	Q9QYR2_MUSMT	Q9qyr2_mus minutol
257	43	69.4	623	1	Q04A4_RABIT	P55787_oryctolagus	330	42	67.7	210	2	Q9QYR4_MUSSA	Q9qyr4_mus saxicol
258	43	69.4	627	2	Q76368_CAEEL	Q76368_caenorhabdi	331	42	67.7	210	2	Q9QYR5_MUSCR	Q9qyr5_mus caroli
259	43	69.4	642	2	Q9BX9_HUMAN	Q9bx9_homo sapien	332	42	67.7	210	2	Q9QYR0_STOLO	Q9qyr0_stochomye 1
260	43	69.4	654	2	Q9BX9_HUMAN	Q9bx9_homo sapien	333	42	67.7	214	1	Q9GR2_MOUSE	Q9gr2_mus musculus
261	43	69.4	666	2	Q9QM05_MOUSE	Q9qmq5_mus musculus	334	42	67.7	218	2	Q5DDN0_SCHJA	Q5ddn0 schistosoma
262	43	69.4	674	1	COA11_BOVIN	P23206_bos taurus	335	42	67.7	222	2	Q7POP7_ANOGA	Q7pop7 anopheles g
263	43	69.4	675	2	Q90800_CHICK	Q90800_gallus gall	336	42	67.7	223	2	Q69BK9_PIG	Q69bk9_sus scrofa
264	43	69.4	708	2	Q7ZWN8_XENLA	Q7zwn8_xenopus lae	337	42	67.7	236	1	COLQ_MOUSE	Q35348_mus musculus
265	43	69.4	712	2	Q4SZ72_TETNG	Q4sz72_tetradodon n	338	42	67.7	241	2	Q4SUE6_TETNG	Q4sue6_tetradodon n
266	43	69.4	747	2	Q6NWE7_BRARE	Q6nw57_brachydanio	339	42	67.7	271	2	Q9VET6_DROME	Q9vet6_drosophila
267	43	69.4	801	2	Q8T5G4_ANOGA	Q8t5g4_anopheles g	340	42	67.7	276	2	Q623J1_CAEBR	Q623j1_caenorhabdi
268	43	69.4	848	2	Q4SZ70_TETNG	Q4sz70_tetradodon n	341	42	67.7	277	2	Q61JU1_CAEBR	Q61ju1_caenorhabdi
269	43	69.4	854	2	Q8IVT9_HUMAN	Q8ivt9_homo sapien	342	42	67.7	282	2	Q61ZNS_CAEBR	Q61zn5_caenorhabdi
270	43	69.4	891	2	Q9OV38_MOUSE	Q9ovs8_mus musculus	343	42	67.7	283	1	YQ33_CAEEL	Q02333_caenorhabdi
271	43	69.4	915	2	Q6ZQK3_MOUSE	Q6zqk3_mus musculus	344	42	67.7	283	2	Q61X02_CAEEL	Q61x02_caenorhabdi
272	43	69.4	962	2	Q61ZNS_CAEBR	Q61zn5_caenorhabdi	345	42	67.7	283	2	Q41989_CAEEL	Q41989_caenorhabdi
273	43	69.4	971	2	Q4RJ71_TETNG	Q4rj71_tetradodon n	346	42	67.7	287	2	Q8CFR0_MOUSE	Q8cfro_mus musculus
274	43	69.4	983	2	Q501R9_RAT	Q501r9_rattus norv	347	42	67.7	291	2	Q19470_CAEEL	Q19470_caenorhabdi
275	43	69.4	988	1	NER1_MOUSE	P97432_mus musculus	348	42	67.7	296	2	Q4RM95_TETNG	Q4rm95_tetradodon n
276	43	69.4	1017	2	Q59HB5_HUMAN	Q59hb5_homo sapien	349	42	67.7	296	2	Q4RM95_TETNG	Q4rm95_tetradodon n
277	43	69.4	1027	2	Q59FB15_HUMAN	Q59fb15_homo sapien	350	42	67.7	299	1	COL34_CAEEL	Q34687_caenorhabdi
278	43	69.4	1139	2	Q5PR22_HUMAN	Q5pr22_homo sapien	351	42	67.7	299	2	Q619V7_CAEBR	Q619v7_caenorhabdi
279	43	69.4	1142	1	COJAI_HUMAN	Q41993_homo sapien	352	42	67.7	299	2	Q61UE9_CAEBR	Q61ue9_caenorhabdi
280	43	69.4	1146	1	COJAI_HUMAN	Q90584_gallus gall	353	42	67.7	300	2	Q61V20_CAEBR	Q61v20_caenorhabdi
281	43	69.4	1347	2	Q8UJ3_ONCKE	Q8uj3_oncorthynch	354	42	67.7	300	2	Q61V22_CAEBR	Q61v22_caenorhabdi
282	43	69.4	1347	2	Q96QB3_HUMAN	Q96qb3_homo sapien	355	42	67.7	302	2	Q19079_CAEEL	Q19079_caenorhabdi
283	43	69.4	1352	2	Q90YU0_BRARE	Q90yjo_brachydanio	356	42	67.7	302	2	Q86NZ7_DROME	Q86nz7_drosophila
284	43	69.4	1352	2	Q61QX2_BRARE	Q61qx2_brachydanio	357	42	67.7	308	2	Q62N27_DROME	Q62n27_drosophila
285	43	69.4	1378	2	Q97405_HALDI	Q97405_haliotis di	358	42	67.7	322	2	Q620L5_CAEBR	Q620l5_caenorhabdi
286	43	69.4	1388	2	Q4RIV6_TETNG	Q4riv6_tetradodon n	359	42	67.7	325	2	Q61QK2_CAEBR	Q61qk2_caenorhabdi
287	43	69.4	1399	2	Q4SB89_TETNG	Q4sb89_tetradodon n	360	42	67.7	329	2	Q21437_CAEEL	Q21437_caenorhabdi
288	43	69.4	1403	2	Q4RKGS_TETNG	Q4rkgs_tetradodon n	361	42	67.7	333	2	FCN1_MOUSE	Q70165_mus musculus
289	43	69.4	1412	2	Q8MUF5_HYDAT	Q8muf5_hydra atten	362	42	67.7	334	1	FCN1_MOUSE	Q70165_mus musculus
290	43	69.4	1445	2	Q93251_RANCA	Q93251_rana cateeb	363	42	67.7	334	2	Q4FJW1_MOUSE	Q4fjw1_mus musculus
291	43	69.4	1447	2	Q91B91_XENLA	Q91b91_xenopus lae	364	42	67.7	335	2	Q60ZK8_CAEBR	Q60zk8_caenorhabdi
292	43	69.4	1447	2	Q6P4U1_BRARE	Q6p4u1_brachydanio	365	42	67.7	368	2	Q61X74_CAEBR	Q61xt4_caenorhabdi
293	43	69.4	1447	2	Q6ULJ5_BRARE	Q6ulj5_brachydanio	366	42	67.7	369	1	SFTPD_BOVIN	Q35246_bos taurus
294	43	69.4	1449	2	Q640B2_XENTR	Q640b2_xenopus tro	367	42	67.7	369	2	Q863A1_BOVIN	Q863a1_bos taurus
295	43	69.4	1449	2	Q802B5_XENLA	Q802b5_xenopus lae	368	42	67.7	370	2	Q02271_CAEEL	Q02271_caenorhabdi
296	43	69.4	1450	2	Q9YIB4_CYNPY	Q9yib4_cynops pyrr	369	42	67.7	370	1	CL46_BOVIN	Q8nhz9_bos taurus
297	43	69.4	1463	2	Q4W6W6_9CHON	Q4w6w6_raja kenofe	370	42	67.7	371	1	CONG_BOVIN	Q23805_bos taurus
298	43	69.4	1486	2	Q71717_XENLA	Q71717_xenopus lae	371	42	67.7	371	2	Q58CU7_BOVIN	Q58cu7_bos taurus
299	43	69.4	1486	2	Q72T16_XENLA	Q72t16_xenopus lae	372	42	67.7	373	2	Q60RN2_CAEBR	Q60rn2_caenorhabdi
300	43	69.4	1496	1	CO5A2_HUMAN	P05997_homo sapien	373	42	67.7	383	2	Q8K2L1_MOUSE	Q8k2l1_mus musculus
301	43	69.4	1496	2	Q53WR4_HUMAN	Q53wr4_homo sapien	374	42	67.7	390	2	Q4SQ42_TETNG	Q4sq42_tetradodon n
302	43	69.4	1497	2	Q61431_MOUSE	Q61431_mus musculus	375	42	67.7	405	2	Q6MWT4_BDEBA	Q6mwt4_bdellovibri
303	43	69.4	1497	2	Q7TMS0_MOUSE	Q7tms0_mus musculus	376	42	67.7	444	2	Q8XCR7_ECO57	Q8xcr7_escherichia
304	43	69.4	1499	2	Q591P2_PIG	Q591p2_sus scrofa	377	42	67.7	448	2	Q5TF53_HUMAN	Q5tf53_homo sapien
305	43	69.4	1502	2	Q59GR4_HUMAN	Q59gr4_homo sapien	378	42	67.7	448	2	Q60SF2_CAEBR	Q60sf2_caenorhabdi
306	43	69.4	1600	2	Q9UEH6_HUMAN	Q9ueh6_homo sapien	379	42	67.7	455	2	Q6DK18_HUMAN	Q6dk18_homo sapien
307	43	69.4	1626	2	Q8NFW1_HUMAN	Q8nfw1_homo sapien	380	42	67.7	456	1	COLQ_HUMAN	Q9v215_homo sapien
308	43	69.4	1631	2	Q597Q0_CANFA	Q597q0_canis famli	381	42	67.7	456	2	Q6YH18_HUMAN	Q6yh18_homo sapien
309	43	69.4	1690	2	Q5JYH8_HUMAN	Q5jyh8_homo sapien	382	42	67.7	456	2	Q6YH19_HUMAN	Q6yh19_homo sapien
310	43	69.4	1691	1	CO4A6_HUMAN	Q14031_homo sapien	383	42	67.7	456	2	Q6YH20_HUMAN	Q6yh20_homo sapien
311	43	69.4	1724	2	Q7PVR8_ANOGA	Q7pvr8_anopheles g	384	42	67.7	456	2	Q6YH21_HUMAN	Q6yh21_homo sapien
312	43	69.4	1747	2	Q26640_STRPU	Q26640_strongyloe	385	42	67.7	458	1	COLQ_RAT	Q35167_rattus norv
313	43	69.4	1758	2	Q61315_CAEEL	Q61315_caenorhabdi	386	42	67.7	486	1	CHST7_HUMAN	Q9ns84_h carbobydr
314	43	69.4	1759	2	Q611Q5_CAEEL	Q611q5_caenorhabdi	387	42	67.7	486	2	Q8NE08_HUMAN	Q8ne08_homo sapien
315	43	69.4	1763	1	CO4A2_ASCSU	P27393_ascaris suu	388	42	67.7	538	2	Q53Q01_HUMAN	Q53q01_homo sapien
316	43	69.4	1820	2	Q91907_PAGMA	Q91907_pagrus majo	389	42	67.7	561	2	Q961F4_HUMAN	Q961f4_homo sapien
317	43	69.4	1827	2	Q589R0_ORYLA	Q589r0_oryzias lat	390	42	67.7	563	2	Q4T6V0_TETNG	Q4t6v0_tetradodon n
318	43	69.4	1835	2	Q8UUM5_ORYLA	Q8uum5_oryzias lat	391	42	67.7	585	2	Q8OV57_MOUSE	Q8ov57_mus musculus
319	43	69.4	1835	2	Q91AU4_CHICK	Q91au4_gallus gall	392	42	67.7	591	2	Q96HCO_HUMAN	Q96hc0_homo sapien
320	43	69.4	2944	1	CO7A1_HUMAN	Q02388_homo sapien	393	42	67.7	662	2	Q8VHY3_MOUSE	Q8vhy3_mus musculus
321	42	67.7	33	3	Q7YOR8_TUPGB	Q7yvr8_tupaia glis	394	42	67.7	670	2	Q4ZA81_9VIRU	Q4za81_bacterioph
322	42	67.7	50	2	Q4RWF7_TETNG	Q4rwf7_tetradodon n	395	42	67.7	671	1	CO1A1_RAT	P02454_rattus norv
323	42	67.7	94	2	O02674_BOVIN	O02674_bos taurus	396	42	67.7	674	2	O7LGG2_HUMAN	Q7lgg2_homo sapien

397	42	67.7	67.7	1	CO9A3_CHICK	P32017	gallus galli
398	42	67.7	67.6	2	O6D850_XENLA	O6des0	xenopus lae
399	42	67.7	679	2	Q6RUD3_HUMAN	Q6RUD3	homo sapien
400	42	67.7	680	2	Q8BSX1_MOUSE	Q8BSX1	mus musculus
401	42	67.7	684	1	CO9A3_HUMAN	CO9A3	homo sapien
402	42	67.7	779	1	CO1A1_BOVIN	P02453	bos taurus
403	42	67.7	809	2	O93485_ONCMY	O93485	oncorhynch
404	42	67.7	847	2	O59F15_HUMAN	O59f15	homo sapien
405	42	67.7	854	2	O09238_9METZ	O09238	pseudocorti
406	42	67.7	890	2	O77087_9ANNE	O77087	alvinella p
407	42	67.7	894	2	O5RC94_PONPY	O5RC94	pongo pygma
408	42	67.7	899	2	O7PVR9_ANOGA	O7PVR9	anopheles g
409	42	67.7	902	2	O16161_MYTED	O16161	mytilus edu
410	42	67.7	905	2	O5B0P8_EMENI	O5B0P8	aspergillus
411	42	67.7	921	1	CO9A1_HUMAN	P20849	homo sapien
412	42	67.7	925	2	O4SIU4_TETNG	O4SIU4	tetraodon n
413	42	67.7	957	2	O641F3_XENLA	O641F3	xenopus lae
414	42	67.7	963	1	BLI1_CABEL	O09457	caenorhabdi
415	42	67.7	966	1	NBR1_HUMAN	O14596	homo sapien
416	42	67.7	966	2	O5J7Q8_HUMAN	O5J7Q8	homo sapien
417	42	67.7	979	2	Q919K3_CHICK	Q919K3	gallus galli
418	42	67.7	1008	2	O5TZ63_BRARE	O5TZ63	brachydanio
419	42	67.7	1019	1	CO6A1_CHICK	P20785	gallus galli
420	42	67.7	1024	2	O8T754_ANOGA	O8T754	anopheles g
421	42	67.7	1045	2	O801S8_XENLA	O801S8	xenopus lae
422	42	67.7	1191	2	O6PCK7_XENLA	O6PCK7	xenopus lae
423	42	67.7	1324	2	O53RW9_HUMAN	Q53RW9	homo sapien
424	42	67.7	1349	2	O8AW17_BRARE	O8AW17	brachydanio
425	42	67.7	1352	2	O5NT95_PAROL	O5NT95	paralichthy
426	42	67.7	1362	1	CO1A2_CHICK	P02467	gallus galli
427	42	67.7	1366	1	CO1A2_CANFA	O46392	canis famil
428	42	67.7	1366	2	Q15177_HUMAN	Q15177	homo sapien
429	42	67.7	1366	2	O7Z5S6_HUMAN	Q7Z5S6	homo sapien
430	42	67.7	1447	2	O5NT96_PAROL	O5NT96	paralichthy
431	42	67.7	1449	2	O910C0_ONCMY	Q910C0	oncorhynch
432	42	67.7	1453	1	CO1A1_CHICK	P02457	gallus galli
433	42	67.7	1468	2	O4S014_TETNG	O4S014	tetraodon n
434	42	67.7	1594	2	Q4RXU2_TETNG	Q4RXU2	tetraodon n
435	42	67.7	1670	1	CO4A3_HUMAN	Q01955	homo sapien
436	42	67.7	1690	1	CO4A4_HUMAN	P53420	homo sapien
437	42	67.7	1690	2	O53WR1_HUMAN	Q53WR1	homo sapien
438	42	67.7	1792	2	O59E87_HUMAN	O59E87	homo sapien
439	42	67.7	1802	2	O17163_BRUMA	O17163	brugia mala
440	42	67.7	1838	1	CO5A1_HUMAN	P20908	homo sapien
441	42	67.7	1838	2	Q15094_HUMAN	Q15094	homo sapien
442	42	67.7	1838	2	O5SUX4_HUMAN	Q5SUX4	homo sapien
443	42	67.7	1838	2	O88207_MOUSE	O88207	mus musculus
444	42	67.7	1840	2	O59IP3_PIG	O59IP3	sus scrofa
445	42	67.7	1840	2	Q60467_CRILO	Q60467	cricetus
446	42	67.7	1840	2	Q9J103_RAT	Q9J103	rattus norv
447	42	67.7	2061	2	O4S2C2_TETNG	Q4S2C2	tetraodon n
448	42	67.7	2551	2	O8CY18_STRR6	O8CY18	streptococc
449	41	66.1	69	2	O4TG58_TETNG	O4TG58	tetraodon n
450	41	66.1	126	2	O4V6M4_DROME	O4V6M4	drosophila
451	41	66.1	141	2	Q4XS80_BOVIN	Q4XS80	bos taurus
452	41	66.1	143	2	Q9QZD3_9MURI	Q9QZD3	rattus ever
453	41	66.1	150	2	O9CZS2_MOUSE	Q9CZS2	mus musculus
454	41	66.1	154	2	O3VEY5_DROME	Q3VEY5	drosophila
455	41	66.1	157	2	O69Y70_HUMAN	O69Y70	homo sapien
456	41	66.1	208	2	O70614_9MURI	O70614	rattus sp.
457	41	66.1	208	2	Q9QYQ9_9MURI	Q9QYQ9	rattus exul
458	41	66.1	210	2	O62172_CABEL	O62172	caenorhabdi
459	41	66.1	210	2	O8HY83_PIG	O8HY83	sus scrofa
460	41	66.1	211	1	PEM_RAT	O62172	caenorhabdi
461	41	66.1	211	2	O4TU77_RAT	O4TU77	rattus norv
462	41	66.1	218	2	O71RG9_HUMAN	O71RG9	homo sapien
463	41	66.1	222	2	Q5CZB5_HUMAN	Q5CZB5	homo sapien
464	41	66.1	228	2	Q810M6_MOUSE	Q810M6	mus musculus
465	41	66.1	232	2	O4RGF4_TETNG	O4RGF4	tetraodon n
466	41	66.1	239	2	O4V811_XENLA	O4V811	xenopus lae
467	41	66.1	240	2	O95M04_BOVIN	O95M04	bos taurus
468	41	66.1	243	1	CIQT5_HUMAN	CIQT5	homo sapien
469	41	66.1	243	1	CIQT5_MOUSE	Q84479	mus musculus

470	41	66.1	243	1	CIQT5_RAT	Q5FVH0	rattus norv
471	41	66.1	243	2	Q8R002_MOUSE	Q8R002	mus musculus
472	41	66.1	243	2	O4ZJN4_MOUSE	Q4ZJN4	mus musculus
473	41	66.1	244	2	O6QWE7_CHICK	O6QWE7	gallus galli
474	41	66.1	244	2	O4VKW2_CHICK	O4VKW2	gallus galli
475	41	66.1	247	1	ADIPO_MOUSE	O60994	mus musculus
476	41	66.1	247	2	O6GTX4_MOUSE	O6GTX4	mus musculus
477	41	66.1	247	2	O8BRW2_MOUSE	O8BRW2	mus musculus
478	41	66.1	247	2	O6GPX5_XENLA	O6GPX5	xenopus lae
479	41	66.1	248	2	O4KMC6_BRARE	O4KMC6	brachydanio
480	41	66.1	249	2	O4SB57_TETNG	O4SB57	tetraodon n
481	41	66.1	253	2	O5TPD7_ANOGA	O5TPD7	anopheles g
482	41	66.1	254	2	O66TM8_9UCUC	O66TM8	monilema a
483	41	66.1	262	2	O4RO89_TETNG	O4RO89	tetraodon n
484	41	66.1	271	2	O9BWP8_HUMAN	O9BWP8	homo sapien
485	41	66.1	272	2	O9DC75_MOUSE	O9DC75	mus musculus
486	41	66.1	276	2	O18536_CABEL	O18536	caenorhabdi
487	41	66.1	281	2	O8BNA3_MOUSE	O8BNA3	mus musculus
488	41	66.1	291	2	O23422_CABEL	O23422	caenorhabdi
489	41	66.1	296	2	O22389_CABEL	O22389	caenorhabdi
490	41	66.1	300	2	O61V23_CABER	O61V23	caenorhabdi
491	41	66.1	300	2	O61V30_CABER	O61V30	caenorhabdi
492	41	66.1	300	2	O86NE3_CABEL	O86NE3	caenorhabdi
493	41	66.1	300	2	O86NE4_CABEL	O86NE4	caenorhabdi
494	41	66.1	301	2	O22379_CABEL	O22379	caenorhabdi
495	41	66.1	301	2	O22380_CABEL	O22380	caenorhabdi
496	41	66.1	301	2	O86FL7_CABEL	O86FL7	caenorhabdi
497	41	66.1	301	2	O9XUT4_CABEL	O9XUT4	caenorhabdi
498	41	66.1	304	2	O19592_CABEL	O19592	caenorhabdi
499	41	66.1	313	2	O9TYB8_HYDAT	O9TYB8	hydra atten
500	41	66.1	321	2	O61TB7_CABER	O61TB7	caenorhabdi

ALIGNMENTS

RESULT 1

Q99228_HUMAN	Q99228_HUMAN PRELIMINARY;	PRT;	584 AA.
ID	Q99228_HUMAN PRELIMINARY;		
AC	Q99228;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DR	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Alpha-1 type XIII collagen.		
GN	Name=COL3A1; (Human)		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;		
OC	Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RX	MEDLINE=91009112; PubMed=1698771;		
RA	Pihlajaniemi T., Tamminen M.;		
RT	"The alpha 1 chain of type XIII collagen consists of three collagenous		
RT	and four noncollagenous domains, and its primary transcript undergoes		
RL	complex alternative splicing.";		
RL	J. Biol. Chem. 265:16922-16928(1990).		
DR	EMBL; M59217; AAA51685.1; -; mRNA.		
DR	Ensembl; ENSG00000197467; Homo sapiens.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0005515; F:protein binding; IEA.		
DR	GO; GO:0005198; F:structural molecule activity; IEA.		
DR	GO; GO:0007155; F:cell adhesion; IEA.		
DR	GO; GO:0006817; P:phosphate transport; IEA.		
DR	InterPro; IPR008161; Clg_helix.		
DR	InterPro; IPR008160; Collagen.		
DR	Pfam; PF01391; Collagen; 8.		
DR	Probdom; PD000007; Clg_helix; 3.		
DR	Collagen.		
SK	SEQUENCE		
QW	SEQUENCE 584 AA; 56522 MW; 090A5BA4F007BF4B CRC64;		
Query Match	93.5%;	Score 58;	DB 2; Length 584;

DR EMBL; AL138925; CAIL15452.1; -; Genomic_DNA.
 DR EMBL; AC024601; CAIL15452.1; JOINED; Genomic_DNA.
 DR EMBL; AC025426; CAIL15452.1; JOINED; Genomic_DNA.
 DR Ensembl; ENSG00000197467; Homo sapiens.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 2.
 KW Collagen.
 SQ SEQUENCE 683 AA; 66377 MW; 3BB550FD03062C30 CRC64;

Query Match 93.5%; Score 58; DB 2; Length 683;
 Best Local Similarity 100.0%; Pred. No. 0.49;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
 |||||
 Db 331 GEKGAGSPGL 341

RESULT 5

Q9NQ52_HUMAN PRELIMINARY; PRT; 717 AA.
 AC Q9NQ52;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Type XIII collagen.
 GN Name=COL13A1; Synonyms=COLXIII1A1;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Pihlajaniemi T.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ293624; CAC00688.1; -; mRNA.
 DR Ensembl; ENSG00000197467; Homo sapiens.
 DR HGNC; HGNC:2190; COL13A1.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.
 DR GO; GO:0005198; F:structural molecule activity; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 717 AA; 69964 MW; A311E9C7D3E87577 CRC64;

Query Match 93.5%; Score 58; DB 2; Length 717;
 Best Local Similarity 100.0%; Pred. No. 0.52;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
 |||||
 Db 353 GEKGAGSPGL 363

RESULT 6

Q5TAT6_HUMAN PRELIMINARY; PRT; 717 AA.
 AC Q5TAT6;
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 DE Collagen, type XIII, alpha 1.

GN Name=COL13A1; ORFNames=RP11-26212.1-001;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Lawlor S.;
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL138925; CAIL15450.1; -; Genomic_DNA.
 DR EMBL; AC024601; CAIL15450.1; JOINED; Genomic_DNA.
 DR EMBL; AC025426; CAIL15450.1; JOINED; Genomic_DNA.
 DR HGNC; HGNC:2190; COL13A1.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 717 AA; 69950 MW; FD12CA80CC93540F CRC64;

Query Match 93.5%; Score 58; DB 2; Length 717;
 Best Local Similarity 100.0%; Pred. No. 0.52;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
 |||||
 Db 353 GEKGAGSPGL 363

RESULT 7

Q8K036_MOUSE PRELIMINARY; PRT; 565 AA.
 ID Q8K036_MOUSE PRELIMINARY;
 AC Q8K036;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Col13a1 protein.
 GN Name=Col13a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=Mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickens M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywicki M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.

```

RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RC Strauberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034164; AAH34164.1; -; mRNA.
DR MGI; MGI:1277201; Coll13a1.
DR GO; GO:0005911; C:intercellular junction; IDA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD000007; C1g_helix; 1.
KW Collagen.
SQ SEQUENCE 565 AA; 56726 MW; DBD3FF99D670195F CRC64;

Query Match 90.3%; Score 56; DB 2; Length 565;
Best Local Similarity 91.7%; Pred. No. 0.89;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GEKAGSGPGLL 12
Db 329 GEKAGSGPGLL 340

RESULT 8
Q8CD80 MOUSE
ID Q8CD80 MOUSE PRELIMINARY; PRT; 568 AA.
AC
DC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
DE enriched library. clone:6030402F20 product:procollagen, type XIII,
DE alpha 1, full insert sequence.
DE Name=Coll13a1;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo I., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;

```

```

RA The PANTOM Consortium.
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo I., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Testis;

```

Query Match 90.3%; Score 56; DB 2; Length 568;
Best Local Similarity 91.7%; Pred. No. 0.89;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GEKAGSGPGLL 12

Db 225 GEKAGSGPGLL 236

RESULT 9

O70575 MOUSE

ID O70575 MOUSE PRELIMINARY; PRT; 739 AA.

AC O70575;

DT 01-AUG-1998 (TrEMBLrel. 07, Created)

DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Collagen type XIII alpha-1 chain.

GN Name=Coll13a1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Gut;
 RX MEDLINE=98288296; PubMed=9624150; DOI=10.1074/jbc.273.25.15590;
 RA Hagg P., Rehn M., Huhtala P., Vaisanen T., Tamminen M.,
 RT "Type XIII collagen is identified as a plasma membrane protein.";
 RL J. Biol. Chem. 273:15590-15597(1998).
 DR EMBL; U30292; AAC24314.1; -; mRNA.
 DR MGI; MGI:1277201; Coll3a1.
 DR GO; GO:0005911; C:intercellular junction; IDA.
 DR InterPro; IPR008161; Clg_helix.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 739 AA; 72110 MW; F13951061381F017 CRC64;

Query Match 90.3%; Score 56; DB 2; Length 739;
 Best Local Similarity 91.7%; Pred. No. 1.2;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGLL 12
 ||||| |||||
 Db 360 GEKGAAGSPGLL 371

RESULT 10

Q9RIN9_MOUSE
 ID Q9RIN9_MOUSE PRELIMINARY; PRT; 751 AA.
 AC Q9RIN9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 GN Type XIII collagen.
 GE Name=Coll3a1; Synonyms=coll3a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=99357014; PubMed=10429945; DOI=10.1016/S0945-053X(99)00018-9;
 RA Kvist A.P., Latvanlehto A., Sund M., Horelli-Kuitunen N., Rehn M.,
 RA Palotie A., Beier D., Pihlajaniemi T.;
 RT "Complete exon-intron organization and chromosomal location of the
 gene for mouse type XIII collagen (coll3a1) and comparison with its
 human homologue.";
 RL Matrix Biol. 18:261-274 (1999).
 RN [2]

Q9RIN9_MOUSE
 ID Q9RIN9_MOUSE PRELIMINARY; PRT; 751 AA.
 AC Q9RIN9;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 GN Type XIII collagen.
 GE Name=Coll3a1; Synonyms=coll3a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=99357014; PubMed=10429945; DOI=10.1016/S0945-053X(99)00018-9;
 RA Kvist A.P., Latvanlehto A., Sund M., Horelli-Kuitunen N., Rehn M.,
 RA Palotie A., Beier D., Pihlajaniemi T.;
 RT "Complete exon-intron organization and chromosomal location of the
 gene for mouse type XIII collagen (coll3a1) and comparison with its
 human homologue.";
 RL Matrix Biol. 18:261-274 (1999).
 RN [2]

DR EMBL; AF063681; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063680; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063679; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063678; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063692; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063691; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063690; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063689; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063688; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063687; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063676; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063674; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063672; AAD50327.1; JOINED; Genomic_DNA.
 DR EMBL; AF063670; AAD50327.1; JOINED; Genomic_DNA.
 DR Ensembl; ENSMUSG0000058806; Mus musculus.
 DR MGI; MGI:1277201; Coll3a1.
 DR GO; GO:0005911; C:intercellular junction; IDA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 9.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen.
 SQ SEQUENCE 751 AA; 73172 MW; FBE2443E1CBF51AD CRC64;

Query Match 90.3%; Score 56; DB 2; Length 751;
 Best Local Similarity 91.7%; Pred. No. 1.2;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGLL 12
 ||||| |||||
 Db 360 GEKGAAGSPGLL 371

RESULT 11

Q16593_HUMAN
 ID Q16593_HUMAN PRELIMINARY; PRT; 447 AA.
 AC Q16593;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
 GN Collagen-like protein (447 AA) (fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Placenta;
 RA Kimura S.;
 RL Submitted (APR-1989) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X14963; CAA33085.1; -; mRNA.
 DR EMBL; X15038; CAA33142.1; -; mRNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR Pfam; PF01391; Collagen; 6.
 DR ProDom; PD000007; Clg_helix; 1.
 KW Collagen.
 FT NON TER 447 447
 SQ SEQUENCE 447 AA; 41829 MW; FDB207023D87CC94 CRC64;

Query Match 83.9%; Score 52; DB 2; Length 447;
 Best Local Similarity 81.8%; Pred. No. 3.3;
 Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGL 11
 ||||| |||||
 Db 416 GEKGAQSPGV 426

```

RT      cysteine-containing non-collagenous segments.";
RL      J. Biochem. 112:856-863(1992).
CC      -1- FUNCTION: The numerous interruptions in the triple helix may make
CC      this molecule either elastic or flexible.
CC      -1- TISSUE SPECIFICITY: In the placenta, where it is found in the
CC      amnion, a membranous tissue lining the amniotic cavity. Within the
CC      complex network of reticular fibers, also located to a fibroblast
CC      layer beneath this dense layer. Exists in tissues in association
CC      with other types of collagen.
CC      -1- DEVELOPMENTAL STAGE: Transiently elevated expression during
CC      gestation, and decrease at term.
CC      -1- DOMAIN: This sequence defines eighteen different domains, nine
CC      triple-helical domains (COL9 to COL1) and ten nontriple-helical
CC      domains (NC10 to NC1).
CC      -1- PTM: Prolines at the third position of the tripeptide repeating
CC      unit (G-X-Y) are hydroxylated in some or all of the chains.
CC      -1- SIMILARITY: Belongs to the fibril-associated collagens with
CC      interrupted helices (FACIT) family.
CC      -1- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC      -----
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use as long as its content is in no way modified and this statement is not
CC      removed.
CC      -----
DR      EMBL; M92642; AAA58427.1; -; mRNA.
DR      EMBL; S57132; AAB25797.1; -; mRNA.
DR      PIR; S23810; S23810.
DR      Ensembl; ENSG00000084636; Homo sapiens.
DR      HGNC; HGNC:2193; COL16A1.
DR      MIM; 120326; -.
DR      GO; GO:0005597; C:collagen type XVI; TAS.
DR      GO; GO:0007565; P:pregnancy; TAS.
DR      InterPro; IPR008161; Clg helix.
DR      InterPro; IPR008160; Collagen.
DR      InterPro; IPR003129; Laminin_G_TSP_N.
DR      Pfam; PF01391; Collagen; 16.
DR      ProDom; PD000007; Clg helix; 1.
DR      SMART; SM00210; TSPN; 1.
KW      Collagen; Extracellular matrix; Hydroxylation; Repeat; Signal;
KW      Structural protein.
FT      SIGNAL 1 21 Potential.
FT      CHAIN 22 1603 Collagen alpha 1(XVI) chain.
FT      DOMAIN 50 231 TSP N-terminal.
FT      REGION 232 374 Nonhelical region 10 (NC10).
FT      REGION 375 505 Triple-helical region 9 (COL9) with 3
FT      REGION 506 520 imperfections.
FT      REGION 521 554 Nonhelical region 9 (NC9).
FT      REGION 555 571 Triple-helical region 8 (COL8) with 1
FT      REGION 572 630 imperfection.
FT      REGION 631 651 Nonhelical region 8 (NC8).
FT      REGION 652 722 Triple-helical region 7 (COL7) with 1
FT      REGION 723 737 imperfection.
FT      REGION 738 875 Nonhelical region 6 (NC6).
FT      REGION 876 886 Triple-helical region 5 (COL5) with 3
FT      REGION 887 938 imperfections.
FT      REGION 939 972 Nonhelical region 5 (NC5).
FT      REGION 973 987 Triple-helical region 4 (COL4) with 2
FT      REGION 988 1010 imperfections.
FT      REGION 1011 1432 Nonhelical region 4 (NC4).
FT      REGION 1433 1471 Triple-helical region 3 (COL3).
FT      REGION 1472 1577 Nonhelical region 3 (NC3).
FT      REGION 1578 1603 Triple-helical region 2 (COL2) with 2
FT      REGION 1603 1603 imperfections.
FT      REGION 1603 1603 Nonhelical region 2 (NC2).
FT      REGION 1603 1603 Triple-helical region 1 (COL1) with 2
FT      REGION 1603 1603 imperfections.
FT      REGION 1603 1603 Nonhelical region 1 (NC1).

RESULT 12
Q59F89 HUMAN PRELIMINARY; PRT; 1463 AA.
AC Q59F89
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Alpha 1 type XVI collagen variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC Tissue=Splice;
RA Totoki Y., Toyoda A., Takeda T., Sakaki Y., Tanaka A., Yokoyama S.,
RA Ohara O., Nagase T., Kikuno F.R.;
RT "None Title.";
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB209571; BAD22808.1; -; mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:000198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 14.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Extracellular matrix.
FT NON_TER 1
FT SEQUENCE 1463 AA; 142798 MW; 48670B10CA93785C CRC64;

Query Match 83.9%; Score 52; DB 2; Length 1463;
Best Local Similarity 81.8%; Pred. No. 11;
Matches .9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEKGAGSGPL 11
|||:||||:
Db 708 GEKGAGSGPGV 718

RESULT 13
COGAL HUMAN STANDARD; PRT; 1603 AA.
AC Q07092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(XVI) chain precursor.
GN Name=COL16A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92335339; PubMed=1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP NUCLEOTIDE SEQUENCE OF 418-1603.
RC Tissue=Placenta;
RX MEDLINE=93203161; PubMed=1284248;
RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
RA Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
chain, alpha 1(XVI), consisting of repetitive collagenous domains and

```

FT CONFLICT 418 420 RDA -> GGR (in Ref. 2).

FT CONFLICT 537 537 R -> P (in Ref. 2).

FT CONFLICT 1160 1160 T -> P (in Ref. 2).

FT CONFLICT 1163 1163 T -> P (in Ref. 2).

FT CONFLICT 1165 1165 S -> P (in Ref. 2).

SQ SEQUENCE 1603 AA; 157693 MW; E27D9A1D4E598A37 CRC64;

Query Match 83.9%; Score 52; DB 1; Length 1603;

Best Local Similarity 81.8%; Pred. No. 13;

Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEKGAEGSPGL 11

DB 817 GEKGAEGSPGV 827

RESULT 14

Q8K388 MOUSE

ID Q8K388 MOUSE PRELIMINARY; PRT; 1046 AA.

AC Q8K388

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Coll16a1 protein.

GN Name=Coll16a1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L.H., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,

RA Bonak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Myers R.M.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska J., Smallos D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RL and mouse cDNA sequences."

RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;

RA Strausberg R.;

RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC027766; AAH27766.1; -; mRNA.

DR MGI; MGI:1095396; Coll16a1.

DR GO; GO:0005737; Cytoplasm; IEA.

DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.

DR GO; GO:0005198; F:structural molecule activity; IEA.

DR GO; GO:0007155; P:cell adhesion; IEA.

DR GO; GO:0006817; P:phosphate transport; IEA.

DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.

DR Pfam; PF01391; Collagen; 5.

DR ProDom; PD000007; C1g_helix; 2.

DR SMART; SM00210; TSPN_1.

KW Collagen; Extracellular matrix; Structural protein.

SQ SEQUENCE 1046 AA; 104656 MW; AD197E2F2F3AF3EF CRC64;

Query Match 82.3%; Score 51; DB 2; Length 1046;

Best Local Similarity 90.0%; Pred. No. 12;

Matches 9; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEKGAEGSPG 10

DB 827 GEKGAEGSPG 836

RESULT 15

Q8BLX7 MOUSE

ID Q8BLX7 MOUSE PRELIMINARY; PRT; 1580 AA.

AC Q8BLX7

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Mus musculus adult male aorta and vein cDNA, RIKEN full-length

DE enriched library, clone:A530052M23 product:COLLAGEN ALPHA 1(XVI) CHAIN

DE homolog

GN Name=Coll16a1;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

OC Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;

RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

RA Carninci P.; Hayashizaki Y.;

RT "High-efficiency full-length cDNA cloning."

RL Meth. Enzymol. 303:19-44 (1999).

RN [2]

NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;

RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;

RA Kawai J., Shinsigawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aizawa K., Izawa K., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Giesi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schraml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."

RL Nature 409:685-690 (2001).

RN [3]

NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;

RA The FANTOM Consortium,

RT "Analysis of the mouse transcriptome based on functional annotation of

RL 60,770 full-length cDNAs."

RN Nature 420:563-573 (2002).

RN [4]

NUCLEOTIDE SEQUENCE.

RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;

RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;

RT "Normalization and subtraction of cap-trapper-selected cDNAs to


```
FT CHAIN 24 689 Collagen alpha 2(IX) chain.
FT REGION 27 163 Triple-helical region 1.
FT REGION 164 180 Nonhelical region 1.
FT REGION 181 519 Triple-helical region 2.
FT REGION 520 549 Nonhelical region 2.
FT REGION 550 632 Triple-helical region 3.
FT REGION 633 634 Nonhelical region 3.
FT REGION 635 664 Triple-helical region 4.
FT REGION 665 689 Nonhelical region 4.
FT CARBOHYD 169 169 O-linked (Xyl.. .) (glycosaminoglycan) (By
FT similarity).
FT DISULFID 174 174 Interchain (Potential).
FT DISULFID 178 178 Interchain (Potential).
FT VARIANT 326 326 Q -> R (in dbSNP:7533552).
FT VARIANT 326 326 Q -> W (in IDB; requires 2 nucleotide
FT substitutions).
FT VARIANT 581 581 V -> I (in dbSNP:3737821).
FT VARIANT 581 581 V -> I (in dbSNP:3737821).
FT SEQUENCE 689 AA; 65131 MW; E86106E02F6FA862 CRC64;

Query Match 80.6%; Score 50; DB 1; Length 689;
Best Local Similarity 81.8%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 277 GKGDEGSPGI 287

RESULT 17
Q619U9 CAEBR PRELIMINARY; PRT; 396 AA.
AC Q619U9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein CBG14092.
GN Name=CBG14092;
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RG The C.briggsae Sequencing Consortium;
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; CAAC0100068; CAB68355.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0043202; F:structural constituent of cuticle; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen_3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 396 AA; 38850 MW; 04EEBD576CC45442 CRC64;

Query Match 79.0%; Score 49; DB 2; Length 396;
Best Local Similarity 81.8%; Pred. No. 9.4;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 265 GKGADGFPGL 275

RESULT 18
Q5C3P1 SCHJA PRELIMINARY; PRT; 518 AA.
ID Q5C3P1 SCHJA PRELIMINARY;
AC Q5C3P1;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
```

```
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
OS Schistosoma japonicum (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatoidea; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6182;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Han Z.;
RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY809845; AAX25734.1; -; mRNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR006108; 3HCDH_C.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR InterPro; IPR011991; Wing_hlx_DNA_bd.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 4.
DR SMART; SM00111; C4; 2.
DR SMART; SM00111; C4; 2.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 518 AA; 52995 MW; 23FD3BE79743B073 CRC64;

Query Match 79.0%; Score 49; DB 2; Length 518;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 12
Db 232 GKGEGQPGLL 243

RESULT 19
O17866 CAEEL PRELIMINARY; PRT; 650 AA.
ID O17866 CAEEL PRELIMINARY;
AC O17866;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE CLE-1C protein (Hypothetical protein cle-1).
GN Name=cle-1; ORFNames=C36B1.1, C36B1.1c;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21157401; PubMed=11257122; DOI=10.1083/jcb.152.6.1219;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,
RA Kramer J.M.;
RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
RT collagen affects cell migration and axon guidance.";
RL J. Cell Biol. 152:1219-1232(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AF164959; AAD47825.1; -; Genomic_DNA.
DR EMBL; Z81079; CAB03084.1; -; Genomic_DNA.
DR PIR; T22002; T22002.
DR HSP; P39061; IKOE.
DR Ensembl; C36B1.1; Caenorhabditis elegans.
DR WormBase; WBGene0000527; cle-1.
DR WormPep; C36B1.1c; CE17816.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0031012; C:extracellular matrix; IEA.
```

Mon Mar 13 10:53:10 2006

DR GO: GO:0005198; F: structural molecule activity; IEA.
 DR GO: GO:0007155; P: cell adhesion; IEA.
 DR GO: GO:0006817; P: phosphate transport; IEA.
 DR InterPro: IPR008160; Collagen.
 DR InterPro: IPR010515; Endostatin.
 DR Pfam: PF01391; Collagen; 2.
 DR Pfam: PF06482; Endostatin; 1.
 KW Collagen; Complete proteome; Hypothetical protein.
 SQ SEQUENCE 650 AA; 69597 MW; 6CF29ED9C16B170E CRC64;

Query Match 79.0%; Score 49; DB 2; Length 650;
 Best Local Similarity 72.7%; Pred. No. 16;
 Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPL 11

Db 206 GEGDAGAPGL 216

RESULT 20

QARGBO_TETNG PRELIMINARY; PRT; 662 AA.
 AC QARGBO;
 DT 13-SEP-2005 (TREMELrel. 31, Created)
 DT 13-SEP-2005 (TREMELrel. 31, Last sequence update)
 DT 13-SEP-2005 (TREMELrel. 31, Last annotation update)
 DE Chromosome 12 SCAP15104, whole genome shotgun sequence.
 DE (Fragment).
 GN ORFNames=GSTENG00034886001;
 OS Tetraodon nigroviridis (Green puffer).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Tetraodon.
 OX NCBI_TaxID=99883;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Jallou O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
 RA Mauceli E., Bounneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
 RA Bismont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
 RA Craud C., Duprat S., Brottier P., Coutanceau J.P., Guzy J.,
 RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
 RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crollius H.,
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
 the early vertebrate proto-karyotype.";
 RL Nature 431:946-957(2004).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RG Genoscope; Whitehead Institute Centre for Genome Research;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBSJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBSJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC -!- SIMILARITY: Contains 1 peptidase M12B domain.
 DR EMBL; CAAG01015104; CAG12572.1; -, Genomic_DNA.
 DR InterPro; IPR006386; ADAM_cysteine.
 DR InterPro; IPR001762; Disintegrin.
 DR InterPro; IPR006209; EGF like.
 DR InterPro; IPR006025; Pept_M_Zn_BS.
 DR InterPro; IPR001590; Peptidase_M12B.
 DR InterPro; IPR002870; Peptidase_M12B_N.
 DR Pfam; PF00200; Disintegrin; 1.
 DR Pfam; PF01562; Pept_M12B_propep; 1.
 DR Pfam; PF01421; Reprolysin; 1.
 DR PRINTS; PR00289; DISINTEGRIN.
 DR ProDom; PD000664; Disintegrin; 1.
 DR SMART; SM00608; ACR; 1.

DR SMART; SM00050; DISIN; 1.
 DR PROSITE; PS02115; ADAM_MEPPO; 1.
 DR PROSITE; PS02114; DISINTEGRIN 2; 1.
 DR PROSITE; PS01186; EGF 2; UNKNOWN 1.
 DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
 FT NON_TER 662
 SQ SEQUENCE 662 AA; 72145 MW; DCA942C9F530BCBD CRC64;

Qy 1 GEGKAGSGPL 12

Db 580 GEGKAGSGPL 591

RESULT 21

COAAL_MOUSE STANDARD; PRT; 680 AA.
 AC COAAL_MOUSE
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 1(X) chain precursor.
 GN Names=Col10a1;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=BALB/c;
 RX MEDLINE=93143676; PubMed=8424763;
 RA Elima K., Eerola I., Rosati R., Metsaranta M., Garofalo S., Perala M.,
 RA de Crombrughe B., Vuorio E.;
 RT "The mouse collagen X gene: complete nucleotide sequence, exon
 structure and expression pattern.";
 RL Biochem. J. 289:247-253(1993).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=129/SV; TISSUE=Liver;
 RX MEDLINE=93238750; PubMed=8477738;
 RA Kong R.Y.C., Kwan K.M., Lau E.T., Thomas J.T., Boot-Handford R.P.,
 RA Grant M.E., Chean K.S.B.;
 RT "Intron-exon structure, alternative use of promoter and expression of
 the mouse collagen X gene, Col10a-1.";
 RL Eur. J. Biochem. 213:99-111(1993).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RX PubMed=8492743;
 RA Apte S.S., Olsen B.R.;
 RT "Characterization of the mouse type X collagen gene.";
 RL Matrix 13:165-179(1993).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 51-680.
 RC STRAIN=DBA/2J;
 RX MEDLINE=92267014; PubMed=1587271;
 RA Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
 RT "Cloning of the human and mouse type X collagen genes and mapping of
 the mouse type X collagen gene to chromosome 10.";
 RL Eur. J. Biochem. 206:217-224(1992).
 RN [5]
 RP NUCLEOTIDE SEQUENCE OF 385-627.
 RC STRAIN=C57BL/6;
 RX MEDLINE=92182017; PubMed=1543751; DOI=10.1016/0167-4781(92)90465-C;
 RA Elima K., Metsaranta M., Kallio J., Perälä M., Eerola I.,
 RA Garofalo S., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse alpha 2(IX) and alpha 1(X)
 collagen mRNAs.";
 RL Biochim. Biophys. Acta 1130:78-80(1992).

CC -1- FUNCTION: Type x collagen is a product of hyperthrophic
CC chondrocytes and has been localized to presumptive
CC mineralization zones of hyaline cartilage.
CC
CC -1- SUBUNIT: Homotrimer.
CC
CC -1- PFM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC
CC -1- SIMILARITY: Contains 1 Clq domain.
CC
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; X67348; CAA47763.1; -; Genomic_DNA.
DR EMBL; X65121; CAA46237.1; -; Genomic_DNA.
DR EMBL; X63013; CAA44741.1; -; mRNA.
DR EMBL; Z21610; CAA79736.1; -; Genomic_DNA.
DR PIR; S31216; S31216.
DR HSP; Q03692; IGR3.
DR SWR; Q05306; 549-680.
DR EMBL; ENSMUSG00000039462; Mus musculus.
DR MGI; MGI:88445; Col10a1.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 6.
DR PRINTS; PR00007; COMPLEMENTC1Q.
DR ProDom; PD000007; Clq_helix; 3.
DR PROSITE; PS00871; Clq; 1.
KW Collagen; Extracellular matrix; Hydroxylation; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 680 Collagen alpha 1(X) chain.
FT DOMAIN 547 680 Clq.
FT REGION 19 56 Nonhelical region (NC2).
FT REGION 57 519 Triple-helical region.
FT REGION 520 680 Nonhelical region (NC1).
FT CONFLICT 13 13 L -> F (in Ref. 3).
FT CONFLICT 27 27 T -> S (in Ref. 3).
FT CONFLICT 248 248 P -> L (in Ref. 3 and 4).
FT CONFLICT 286 286 A -> S (in Ref. 2).
FT CONFLICT 306 306 L -> F (in Ref. 3 and 4).
FT CONFLICT 417 417 R -> S (in Ref. 3 and 4).
FT CONFLICT 451 451 R -> K (in Ref. 5).
FT CONFLICT 500 500 H -> L (in Ref. 3 and 4).
FT CONFLICT 567 572 APIPPD -> CPHPY (in Ref. 3 and 4).
FT CONFLICT 635 635 Q -> T (in Ref. 3 and 4).
SQ SEQUENCE 680 AA; 66775 MW; F5984CA99AF708E2 CRC64;
Query Match 79.0%; Score 49; DB 1; Length 680;
Best Local Similarity 72.7%; Pred. No. 17;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEKAGSGPGL 11
Db 174 GEKAGGAPGV 184
RESULT 22
Q9U9K6 CAEEL PRELIMINARY; PRT; 778 AA.
AC Q9U9K6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE CLP-1B protein (Hypothetical protein clp-1).
GN Name=clp-1; ORFNames=C36B1.1, C36B1.1B;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
CC

OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21157401; PubMed=11257122; DOI=10.1083/jcb.152.6.1219;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,
RA Kramer J.M.;
RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
RT collagen affects cell migration and axon guidance.";
RL J. Cell Biol. 152:1219-1232(2001).
[2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RP STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG "The C. elegans sequencing consortium;
RT investigating biology.";
RL Science 282:2012-2018 (1998).
DR EMBL; AF164959; AAD47824.1; -; Genomic_DNA.
DR EMBL; Z80215; CAD21700.1; -; Genomic_DNA.
DR EMBL; Z81079; CAD21643.1; -; Genomic_DNA.
DR EMBL; Z81079; CAD21700.1; JOINED; Genomic_DNA.
DR EMBL; Z80215; CAD21643.1; JOINED; Genomic_DNA.
DR HSP; P39061; 1KOE.
DR Ensembl; C36B1.1; Caenorhabditis elegans.
DR WormBase; WBGene0000527; clp-1.
DR WormPep; C36B1.1b; CE29706.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005578; C:extracellular matrix (sensu Metazoa); IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF06482; Endostatin; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Complete proteome; Extracellular matrix;
KW Hypothetical protein; Structural protein.
SQ SEQUENCE 778 AA; 83788 MW; 743A7FCA7FF1ABD CRC64;
Query Match 79.0%; Score 49; DB 2; Length 778;
Best Local Similarity 72.7%; Pred. No. 19;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEKAGSGPGL 11
Db 334 GERGAGAPGL 344
RESULT 23
Q9U9K7 CAEEL PRELIMINARY; PRT; 1117 AA.
AC Q9U9K7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CLP-1A protein.
GN Name=clp-1; ORFNames=C36B1.1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21157401; PubMed=11257122; DOI=10.1083/jcb.152.6.1219;
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,
RA Kramer J.M.;
RT "The NCI/endostatin domain of Caenorhabditis elegans type XVIII
RT collagen affects cell migration and axon guidance.";
RL J. Cell Biol. 152:1219-1232(2001).
DR EMBL; AF164959; AAD47823.1; -; Genomic_DNA.
DR HSP; P39061; 1KOE.
DR

DR Ensembl; C36B1.1; Caenorhabditis elegans.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0031012; C:extracellular matrix; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR003129; TSP N.
DR Pfam; PF01391; Collagen_2.
DR Pfam; PF06482; Endostatin; 1.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00853; FN3; 2.
SQ SEQUENCE 1117 AA; 121680 MW; CF323D42E88770AD CRC64;

Query Match 79.0%; Score 49; DB 2; Length 1117;
Best Local Similarity 72.7%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 673 GERGADGAPGL 683
|||:|:|:|
|:|:|:|:|

RESULT 24
ID Q7JL30 CAEEL PRELIMINARY; PRT; 1138 AA.
AC Q7JL30.
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein cle-1.
GN Name=cle-1; ORFNames=C36B1.1, C36B1.1a;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81079; CAD21644.2; -, Genomic DNA.
DR EMBL; Z80215; CAD21701.2; -, Genomic DNA.
DR EMBL; Z80215; CAD21644.2; JOINED; Genomic DNA.
DR EMBL; Z81079; CAD21701.2; JOINED; Genomic DNA.
DR HSSP; P39061; 1KOE.
DR Ensembl; C36B1.1; Caenorhabditis elegans.
DR WormBase; WBGene0000527; cle-1.
DR WormPep; C36B1.1a; C362988.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0031012; C:extracellular matrix; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF06482; Endostatin; 1.
DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00853; FN3; 2.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 1138 AA; 124178 MW; B2C69581A4BCFB80 CRC64;

Query Match 79.0%; Score 49; DB 2; Length 1175;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 731 GERGADGAPGL 741
|||:|:|:|
|:|:|:|:|

RESULT 26
ID Q5VZA9 HUMAN PRELIMINARY; PRT; 1706 AA.
AC Q5VZA9.
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

Query Match 79.0%; Score 49; DB 2; Length 1138;
Best Local Similarity 72.7%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 694 GERGADGAPGL 704
|||:|:|:|
|:|:|:|:|

RESULT 25
ID Q810G6 CAEEL PRELIMINARY; PRT; 1175 AA.
AC Q810G6.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein cle-1.
GN Name=cle-1; ORFNames=C36B1.1, C36B1.1d;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=99069613; PubMed=9851916;
RG The C. elegans sequencing consortium;
RT "Genome sequence of the nematode C. elegans: a platform for investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z80215; CAD56563.2; -, Genomic DNA.
DR EMBL; Z81079; CAD56581.2; -, Genomic DNA.
DR EMBL; Z81079; CAD56563.2; JOINED; Genomic DNA.
DR EMBL; Z80215; CAD56581.2; JOINED; Genomic DNA.
DR HSSP; P39061; 1KOE.
DR Ensembl; C36B1.1; Caenorhabditis elegans.
DR WormBase; WBGene0000527; cle-1.
DR WormPep; C36B1.1d; C37210.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0031012; C:extracellular matrix; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003961; FN_III.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF06482; Endostatin; 1.
DR ProDom; PD000007; Clg_helix; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS00853; FN3; 2.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 1175 AA; 128422 MW; E3E1D2F56C652455 CRC64;

Query Match 79.0%; Score 49; DB 2; Length 1175;
Best Local Similarity 72.7%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 731 GERGADGAPGL 741
|||:|:~|:|
|:|:|:|:|

RESULT 26
ID Q5VZA9 HUMAN PRELIMINARY; PRT; 1706 AA.
AC Q5VZA9.
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

```

DE Collagen, type IV, alpha 2.
GN Name=COL4A2; ORFNames=RP11-90L1.2-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Peck A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Pearce A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Tracey A.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161773; CAH71366.1; -; Genomic DNA.
DR EMBL; AL159153; CAH72050.1; -; Genomic DNA.
DR EMBL; AL139385; CAH17005.1; -; Genomic DNA.
DR EMBL; AL139385; CAH71366.1; JOINED; Genomic DNA.
DR EMBL; AL159153; CAH71366.1; JOINED; Genomic DNA.
DR EMBL; AL159153; CAH17005.1; JOINED; Genomic DNA.
DR EMBL; AL161773; CAH72050.1; JOINED; Genomic DNA.
DR EMBL; AL139385; CAH72050.1; JOINED; Genomic DNA.
DR EMBL; AL161773; CAH17005.1; JOINED; Genomic DNA.
DR SRR; Q5V2A9; 1481-1705.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 20.
DR ProDom; PD000007; Clg_helix; 12.
DR ProDom; PD003923; Procollagn4_C; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE. 1706 AA; 167158 MW; 6PCE70B51F06881 CRC64;

Query Match 79.0%; Score 49; DB 2; Length 1706;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GEKAGSGPL 11
    |||||
Db 270 GEKSGEFGI 280

RESULT 27
CO4A2 HUMAN STANDARD; PRT; 1712 AA.
ID CO4A2 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Collagen alpha 2 (IV) chain precursor.
GN Name=COL4A2;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=83066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV collagen and comparison with the alpha 1 (IV) chain.";

J. Biol. Chem. 263:19488-19493 (1988).
[2]
RN NUCLEOTIDE SEQUENCE OF 1-1042.
RC TISSUE=Placenta;
RX MEDLINE=88151998; PubMed=3345760;
RA Brazel D., Pollner R., Oberbauer I., Kuehn K.;
RT "Human basement membrane collagen (type IV). The amino acid sequence of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain reveals deletions in the alpha 1(IV) chain.";
RL Eur. J. Biochem. 172:35-42 (1988).
[3]
RN NUCLEOTIDE SEQUENCE OF 1254-1712.
RX MEDLINE=87219158; PubMed=3582677; DOI=10.1016/0014-5793 (87) 80706-8;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2 chain cDNA reveals extensive homology with the NC-1 domain of alpha 1 (IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286 (1987).
[4]
RN NUCLEOTIDE SEQUENCE OF 1451-1485.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1 (IV) and alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516 (1987).
[5]
RN NUCLEOTIDE SEQUENCE OF 1486-1712.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238 (1987).
[6]
RN NUCLEOTIDE SEQUENCE OF 1-33.
RX MEDLINE=89034231; PubMed=3182844;
RA Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220 (1988).
[7]
RN NUCLEOTIDE SEQUENCE OF 1-33.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure.";
RL EMBO J. 7:2687-2695 (1988).
[8]
RN NUCLEOTIDE SEQUENCE OF 1-33.
RC TISSUE=Skin;
RX MEDLINE=93305049; PubMed=8317999;
RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription.";
RL Biochem. J. 292:687-695 (1993).
[9]
RN PROTEIN SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624 (1988).
[10]
RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1485-1712.
RX MEDLINE=2208067; PubMed=12011424; DOI=10.1073/pnas.062183499;
RA Than M.E., Henrich S., Huber R., Ries A., Mann K., Kuhn K., Timpl R., Bourenkov G.P., Bartunik H.D., Bode W.;
RT "The 1.9-A crystal structure of the noncollagenous (NC1) domain of human placenta collagen IV shows stabilization via a novel type of

```

RT covalent Met-Lys cross-link.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:6607-6612(2002).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and
CC entactin/nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- PTM: The trimeric structure of the NC1 domains may be stabilized
CC by covalent bonds between Lys and Met residues.
CC -!- SIMILARITY: Belongs to the type IV collagen family.
CC -!- SIMILARITY: Contains 2 COL4C (collagen IV C-terminal) domains.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified, and this statement is not
CC removed.
CC -----
DR EMBL; X05562; CAA29076.1; -; mRNA.
DR EMBL; X05610; CAA29098.1; -; mRNA.
DR EMBL; J02760; AAB58422.1; -; mRNA.
DR EMBL; M36963; AAB53099.1; -; Genomic_DNA.
DR EMBL; X12784; CAA31275.1; -; Genomic_DNA.
DR EMBL; J04217; AAB53097.1; -; Genomic_DNA.
DR PIR; A32024; CGH02B.
DR PDB; 1L11; X-ray; C/F=1485-1712.
DR HGNC; HGNC:2203; COL4A2.
DR MIM; 120090; -.
DR GO; GO:000587; C:collagen type IV; TAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0030198; P:extracellular matrix organization and biogenesis; TAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagen4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 20.
DR ProDom; PD000007; Clg helix; 11.
DR ProDom; PD003923; Procollagen4_C; 2.
DR SMART; SM00111; C4; 2.
DR 3D-structure; Basement membrane; Collagen; Direct protein sequencing;
KW Extracellular matrix; Glycoprotein; Hydroxylation; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 25
FT PROPEP 26 193
FT CHAIN 184 1712
FT DOMAIN 1489 1596
FT REGION 1597 1712
FT REGION 184 1484
FT CARBOHYD 138 138
FT DISULFID 1504 1593
FT DISULFID 1537 1590
FT DISULFID 1549 1555
FT DISULFID 1612 1708
FT DISULFID 1646 1705
FT DISULFID 1658 1665
FT CONFLICT 471 471
FT CONFLICT 683 683
FT CONFLICT 1575 1575
FT CONFLICT 1663 1663
FT CONFLICT 1701 1701
R -> P (in Ref. 2).
A -> G (in Ref. 2).
M -> I (in Ref. 5).
G -> H (in Ref. 9).
H -> G (in Ref. 9).

SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
Query Match 79.0%; Score 49; DB 1; Length 1712;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 GEGAGGSPGL 11
Db 276 GEGSGEGPGI 286
RESULT 28
COB2 MOUSE
ID COB2_MOUSE STANDARD; PRT; 1736 AA.
AC Q64739; Q61432; Q921W0;
DT 15-JUL-1998 (Rel. 36, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN Name=Colla2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridea; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP NUCLEOTIDE SEQUENCE (ISOFORM 7).
RC STRAIN=129/SVJ;
RA Rowen L., Qin S., Madan A., Loretz C., James R., Dors M., Mix L.,
RA Hall J., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class II
RT region.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP NUCLEOTIDE SEQUENCE OF 1-1678 (ISOFORM 7).
RC STRAIN=129/SV, and FVB/N; TISSUE=Cartilage.
RX MEDLINE=97135795; PubMed=898332; DOI=10.1016/S0945-053X(96)90139-0;
RA Vandenbergh P., Vuoristo M.M., Ala-Kokko L., Prockop D.J.;
RT "The mouse colla2 gene. Some transcripts from the adjacent rxr-beta
RT gene extend into the colla2 gene.";
RL Matrix Biol. 15:359-367(1996).
RN (3)
RP NUCLEOTIDE SEQUENCE OF 1-624 (ISOFORMS 1; 2; 3; 4; 5 AND 6).
RC STRAIN=129/SV;
RX MEDLINE=95138212; PubMed=7836472; DOI=10.1074/jbc.270.5.2372;
RA Tsumaki N., Kimura T.;
RT "Differential expression of an acidic domain in the amino-terminal
RT propeptide of mouse pro-alpha2(XI) collagen by complex alternative
RT splicing.";
RL J. Biol. Chem. 270:2372-2378(1995).
RN (4)
RP NUCLEOTIDE SEQUENCE OF 1-8.
RC STRAIN=129/SV; TISSUE=Liver.
RX MEDLINE=96427460; PubMed=8830784; DOI=10.1083/jcb.134.6.1573;
RA Tsumaki N., Kimura T., Matsui Y., Ochi T.;
RT "Separable cis-regulatory elements that contribute to tissue- and
RT site-specific alpha 2(XI) collagen gene expression in the embryonic
RT mouse cartilage.";
RL J. Cell Biol. 134:1573-1582(1996).
CC -!- FUNCTION: May play an important role in fibrillogenesis by
CC controlling lateral growth of collagen II fibrils (By similarity).
CC -!- SUBUNIT: Trimer composed of three different chains: alpha 1(XI),
CC alpha 2(XI), and alpha 3(XI). Alpha 3(XI) is a post-translational
CC modification of alpha 1(XI). Alpha 1(XI) can also be found instead
CC of alpha 3(XI)=1(XI) (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=7;
CC Comment=Additional isoforms seem to exist;
CC Name=1; Synonyms=E56789;
CC IsoId=Q64739-2; Sequence=Displayed;
CC Name=2; Synonyms=E5689;
CC IsoId=Q64739-3; Sequence=VSP_007346;
CC Name=3; Synonyms=E5789;

"Genetic correction of canine dystrophic epidermolysis bullosa mediated by retroviral vectors.";
Hum. Mol. Genet. 12:1897-1905(2003).
-!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.
EMBL; AY183408; AAO64414.1; -, mRNA.
HSSP; P00981; 1DPK.
DR Ensembl; ENSCAFG00000012403; Canis familiaris.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002223; Prot_inh_Kunz-m.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01391; Collagen; 26.
DR Pfam; PF00041; fn3; 9.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00092; VWA; 2.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD000007; Clg helix; 10.
DR ProDom; PD000222; Prot_inh_Kunz-m; 1.
DR SMART; SMO0060; FN3; 9.
DR SMART; SMO0327; VWA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS00853; FN3; 9.
DR PROSITE; PS0234; VWA; 2.
DR Collagen; Extracellular matrix; Structural protein.
SQ SEQUENCE 2936 AA; 293983 MW; F80CAPA0E1F69989 CRC64;
Query Match 79.0%; Score 49; DB 2; Length 2936;
Best Local Similarity 81.8%; Pred. No. 76;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 GEGKAGSGPGL 11
DB 1871 GEGKAGSGPGL 1881
RESULT 31
Q25617 ONCVO PRELIMINARY; PRT; 60 AA.
AC Q25617
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collagen (Fragment).
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96240155; PubMed=8644904;
RA Garate T., Contrachs F.J., Harnett W., Buttner D.W., Parkhouse R.M.;
RT "Identification of Onchocerca volvulus collagen as an antigen mainly recognized by antibodies in chronic hyper-reactive onchodermatitis (sowda).";
RT Am. J. Trop. Med. Hyg. 54:490-497(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Bonay P.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89450; CAA61620.1; -, mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 1.
FT NON_TER 1 1
FT NON_TER 60 60

RT "Genetic correction of canine dystrophic epidermolysis bullosa mediated by retroviral vectors.";
Hum. Mol. Genet. 12:1897-1905(2003).
-!- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.
EMBL; AY183408; AAO64414.1; -, mRNA.
HSSP; P00981; 1DPK.
DR Ensembl; ENSCAFG00000012403; Canis familiaris.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0004867; F:serine-type endopeptidase inhibitor activity; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR002223; Prot_inh_Kunz-m.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01391; Collagen; 26.
DR Pfam; PF00041; fn3; 9.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00092; VWA; 2.
DR PRINTS; PR00759; BASICPTASE.
DR PRINTS; PR00453; VWFADOMAIN.
DR ProDom; PD000007; Clg helix; 10.
DR ProDom; PD000222; Prot_inh_Kunz-m; 1.
DR SMART; SMO0060; FN3; 9.
DR SMART; SMO0327; VWA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS0279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS00853; FN3; 9.
DR PROSITE; PS0234; VWA; 2.
DR Collagen; Extracellular matrix; Structural protein.
SQ SEQUENCE 2936 AA; 293983 MW; F80CAPA0E1F69989 CRC64;
Query Match 79.0%; Score 49; DB 2; Length 2936;
Best Local Similarity 81.8%; Pred. No. 76;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 GEGKAGSGPGL 11
DB 1871 GEGKAGSGPGL 1881
RESULT 31
Q25617 ONCVO PRELIMINARY; PRT; 60 AA.
AC Q25617
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collagen (Fragment).
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=96240155; PubMed=8644904;
RA Garate T., Contrachs F.J., Harnett W., Buttner D.W., Parkhouse R.M.;
RT "Identification of Onchocerca volvulus collagen as an antigen mainly recognized by antibodies in chronic hyper-reactive onchodermatitis (sowda).";
RT Am. J. Trop. Med. Hyg. 54:490-497(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Bonay P.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; X89450; CAA61620.1; -, mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 1.
FT NON_TER 1 1
FT NON_TER 60 60

SQ SEQUENCE 60 AA; 5876 MW; 6290EF720761D8D6 CRC64;
Query Match 77.4%; Score 48; DB 2; Length 60;
Best Local Similarity 66.7%; Pred. No. 1.9;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
OY 1 GEGKAGSGPGL 12
DB 39 GEGKAGSGPGL 50
RESULT 32
P78429 HUMAN PRELIMINARY; PRT; 132 AA.
AC P78429;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 (III) collagen (aa 892-1023) (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
DR EMBL; X01655; CAA25821.1; -, mRNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 2.
FT NON_TER 1 1
FT NON_TER 132 132
SQ SEQUENCE 132 AA; 11836 MW; 2DFCD73FCB6866EE CRC64;
Query Match 77.4%; Score 48; DB 2; Length 132;
Best Local Similarity 80.0%; Pred. No. 4.4;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 1 GEGKAGSGPGL 10
DB 106 GEGKAGSGPGL 115
RESULT 33
Q8BJU6 MOUSE PRELIMINARY; PRT; 338 AA.
AC Q8BJU6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Mus musculus 12 days embryo embryonic body between diaphragm region and neck cDNA, RIKEN full-length enriched library, clone:9430010K02 product:COLLAGEN ALPHA 1(III) CHAIN, full insert sequence.
DE (Fragment).
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J;
RC TISSUE=Embryonic body between diaphragm region and neck;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;

[6] PROTEIN SEQUENCE OF 948-1049.
RA MEDLINE=9005603; PubMed=488911;
RA Almann H., Fiezek P.P., Glanville R.W., Kuhn K.;
RT "The covalent structure of calf skin type III collagen. VI. The amino
RT acid sequence of the carboxyterminal cyanogen bromide peptide alpha
RT 1(III)CB9g (positions 928-1028).";
RL Hoppe-Sewler's Z. Physiol. Chem. 360:861-868(1979).
CC -1- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -1- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylsines.
CC -1- PM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- SIMILARITY: Belongs to the fibrillar collagen family.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC PIR; A02862; CGB07S.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; Clg_helix; 4.
DR PROSITE; PS01208; VWF_1; PARTIAL.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Hydroxylation; Repeat; Structural protein.
FT REGION 1 14 Nonhelical region (N-terminal).
FT REGION 15 1040 Triple-helical region (C-terminal).
FT REGION 1041 1049 5-hydroxyllysine.
FT MOD_RES 95 95 5-hydroxyllysine.
FT MOD_RES 107 107 5-hydroxyllysine.
FT MOD_RES 119 119 5-hydroxyllysine.
FT MOD_RES 938 938 5-hydroxyllysine.
FT MOD_RES 950 950 5-hydroxyllysine.
FT CARBOHYD 107 107 O-linked (Gal. . .).
FT DISULFID 950 950 O-linked (Gal. . .).
FT DISULFID 1040 1040 Interchain.
FT DISULFID 1041 1041 Interchain.
SQ SEQUENCE 1049 AA; 93651 MW; 8EEC33D1C66EC9A3 CRC64;

Query Match 77.4%; Score 48; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGAGSGSPG 10
Db 1014 GERGSESGSPG 1023
||:|||||

RESULT 37
QBN6U4_HUMAN
ID QBN6U4_HUMAN PRELIMINARY; PRT; 1163 AA.
AC QBN6U4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE COL3A1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Warra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028178; AAH28178.1; -; mRNA.
DR Ensembl; ENSG00000168542; Homo sapiens.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI_1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_2; 1.
KW Collagen.
SQ SEQUENCE 1163 AA; 111898 MW; 9E0C6BE1E94D6357 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1163;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGAGSGSPG 10
Db 867 GERGSESGSPG 876
||:|||||

RESULT 38
Q8K173_MOUSE
ID Q8K173_MOUSE PRELIMINARY; PRT; 1222 AA.
AC Q8K173;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Col3a1 protein (Fragment).
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RX TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Mix FVB/N;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028248; AAH28248.1; -; mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 16.
DR ProDom; PD0002078; Fib collagen_C; 1.
DR ProDom; PD0002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR Collagen.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 1222 AA; 115140 MW; A409CA00D82765E4 CRC64;
Query Match 77.4%; Score 48; DB 2; Length 1222;
Best Local Similarity 80.0%; Pred. No. 45;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGGEGSPG 10
Db 927 GEGGEGSPG 936
RESULT 39
QWAW11 BRARE
ID QWAW11 BRARE PRELIMINARY; PRT; 1258 AA.
AC QWAW11;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE SI:BY143E18.1 (Novel protein similar to human alpha 2 type V collagen
DE (COL5A2)) (fragment).
GN Name=col5a21; Synonyms=SI:BY143E18.1;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC Clark G.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL731564; CAD58730.1; -; Genomic_DNA.

DR Ensembl; ENSDARG0000024847; Danio rerio.
DR ZFIN; ZDB-GENE-030616-13; col5a21.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0005515; P:protein binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 15.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg helix; 8.
DR ProDom; PD000007; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
FT NON TER 1
SQ SEQUENCE 1258 AA; 122824 MW; 934928F4E462B817 CRC64;
Query Match 77.4%; Score 48; DB 2; Length 1258;
Best Local Similarity 80.0%; Pred. No. 46;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEGGEGSPG 10
Db 543 GEGGEGSPG 552
RESULT 40
QSPQT6 RAT
ID QSPQT6 RAT PRELIMINARY; PRT; 1463 AA.
AC QSPQT6;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Procollagen, type III, alpha 1.
GN Name=Col3a1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahney J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]

RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Lung; Project;
RA Director MGC Project;
RL Submitted (DEC-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC087039; AAH87039.1; -, mRNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; F:phosphate transport; IEA.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLF1; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; C1g_helix; 3.
DR ProDom; PD02078; Fib_collagen_C; 1.
DR SMART; SM00038; COLF1_1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1463 AA; 138936 MW; 63C218CD2BCA47B6 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1463;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GEGKAGSGSPG 10
DB 1169 GERGEGSGPG 1178
|||:|||||
|||:|||||

RESULT 41
CO3AL MOUSE STANDARD; PRT; 1464 AA.
AC P08121; Q61429; Q9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
GN [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795; DOI=10.1016/0378-1119(94)90061-2;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RT complete DNA sequence."
RL Gene 147:161-168(1994).
RN [2]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RP STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=23388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heaton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RN NUCLEOTIDE SEQUENCE OF 1-488.
RP MEDLINE=88167858; PubMed=3443309; DOI=10.1016/0378-1119(87)90117-X;
RX Wood L., Theriault N., Vogeli G.;
RA "Complete nucleotide sequence of the N-terminal domains of the murine
RT alpha-1 type-III collagen chain."
RL Gene 61:225-230(1987).
RN [4]
RN NUCLEOTIDE SEQUENCE OF 1-28.
RP MEDLINE=85131189; PubMed=3972847;
RX Liaw G., Mudryj M., de Crombrughe B.;
RA "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene."
RL J. Biol. Chem. 260:3773-3777(1985).
RN [5]
RN NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA] OF 810-1464.
RP STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazier K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang L., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
RN [6]
RN NUCLEOTIDE SEQUENCE OF 1442-1464.
RP STRAIN=C57BL/6;
RX MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs."
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -!- FUNCTION: Collagen type III occurs in most soft connective tissues
CC along with type I collagen.
CC -!- SUBUNIT: Trimers of identical alpha 1(III) chains. The chains are
CC linked to each other by interchain disulfide bonds. Trimers are
CC also cross-linked via hydroxylsines.
CC -!- PTM: Proline residues at the third position of the tripeptide
CC repeating unit (G-X-Y) are hydroxylated in some or all of the
CC chains.
CC -!- PTM: O-linked glycan consists of a Glc-Gal disaccharide bound to
CC the oxygen atom of a post-translationally added hydroxyl group (By
CC similarity).


```

CC CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC CC -!- SIMILARITY: Contains 1 WFCC domain.
CC CC -----
CC CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use as long as its content is in no way modified and this statement is not
CC CC removed.
CC CC -----
DR DR EMBL; X52046; CA36279.1; -; Genomic_DNA.
DR DR EMBL; BC043089; AA43089.1; -; mRNA.
DR DR EMBL; BC058724; AA58724.1; -; mRNA.
DR DR EMBL; M18933; AA37338.1; -; mRNA.
DR DR EMBL; K03037; -; NOT ANNOTATED CDS; Genomic_DNA.
DR DR EMBL; AK019448; BAB31724.1; -; mRNA.
DR DR EMBL; X57983; CA441048.1; -; Genomic_DNA.
DR DR PIR; A27353; A27353.
DR DR PIR; S59856; S59856.
DR DR Ensembl; ENSMUSG0000001506; Mus musculus.
DR DR MGI; MGI:88453; Col3a1.
DR DR InterPro; IPR008161; C1g_helix.
DR DR InterPro; IPR008160; Collagen.
DR DR InterPro; IPR000885; Fib_collagen_C.
DR DR InterPro; IPR001007; WFC.
DR DR Pfam; PF01410; COLF1; 1.
DR DR Pfam; PF01391; Collagen; 17.
DR DR ProDom; PD000007; C1g_helix; 3.
DR DR ProDom; PD002078; Fib_collagen_C; 1.
DR DR PROSITE; PS01208; WFC1; 1.
DR DR PROSITE; PS0184; WFC2; 1.
KW KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation; Repeat;
KW Signal; Structural protein.
FT FT SIGNAL 1 23 By similarity.
FT FT PROPEP 24 154 N-terminal propeptide.
FT FT CHAIN 155 1203 Collagen alpha 1(III) chain.
FT FT PROPEP 1204 1464 C-terminal propeptide.
FT FT DOMAIN 31 90 WFC.
FT FT REGION 155 169 Nonhelical region (N-terminal).
FT FT REGION 170 1195 Triple-helical region.
FT FT REGION 1196 1464 Nonhelical region (C-terminal).
FT FT MOD_RES 262 262 5-hydroxylysine (By similarity).
FT FT MOD_RES 283 283 5-hydroxylysine (By similarity).
FT FT MOD_RES 859 859 5-hydroxylysine (By similarity).
FT FT MOD_RES 976 976 5-hydroxylysine (By similarity).
FT FT MOD_RES 1093 1093 5-hydroxylysine (By similarity).
FT FT MOD_RES 1105 1105 5-hydroxylysine (By similarity).
FT FT CARBOHYD 262 262 O-linked (Gal..) (By similarity).
FT FT DISULFID 1195 1195 Interchain (By similarity).
FT FT DISULFID 1196 1196 Interchain (By similarity).
FT FT SEQUENCE 1464 AA; 138943 MW; 2104EC27A86090B CRC64;
Query Match 77.4%; Score 48; DB 1; Length 1464;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 GSKGAEQSPG 10
DB 1169 GERSEGSFG 1178
|||||
|||||
RESULT 42
QBKY2 MOUSE PRELIMINARY; PRT; 1464 AA.
AC QBKY2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:C130072F01 product:COLLAGEN ALPHA 1(III) CHAIN, full
DE insert sequence.
GN Name-Col3a1;
OS Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC OC Muridae; Murinae; Mus.
OC OC NCBI TaxID=10090;
RN RN [1]
RP NUCLEOTIDE SEQUENCE
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=92279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuiki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 403:685-690(2001).
RN RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multiplexed sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN RN [6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Konno S., Konno H., Kouda M., Koya S.,

```

```

RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,
RA Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK048546; BAC33370.1; -; mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VVFC_1; UNKNOWN_1.
DR PROSITE; PS0184; VVFC_2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138946 MW; 1E4ED9539FE42C12 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1464;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
Dd 1169 GERGSEGPSG 1178

RESULT 43
Q8BLW4 MOUSE
ID Q8BLW4 MOUSE PRELIMINARY; PRT; 1464 AA.
AC Q8BLW4.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Mus musculus adult male aorta and vein cDNA, RIKEN full-length
DE enriched library, clone:A530083N24 product:COLLAGEN ALPHA 1(III)
DE CHAIN, full insert sequence.
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staab R., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Harsh G.,

```

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA The PANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.

```

```

DR PROSITE, PS01208; VWF_C1; UNKNOWN 1.
DR PROSITE, PS0184; VWF_C2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138938 MW; 91F3246D90818449 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1464;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|:|||||
Db 1169 GERGSEGSPPG 1178

RESULT 44
Q7TT32 MOUSE
ID Q7TT32_MOUSE PRELIMINARY; PRT; 1464 AA.
AC Q7TT32;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Procollagen, type III, alpha 1.
GN Name=Col3a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
EX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Hellon E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC052398; AAHS2398.1; -; mRNA.
DR Ensembl; ENSMUSG0000001506; Mus musculus.
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005733; C:cytosol; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VMC; 1.
DR ProDom; PD000007; C1g helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.

DR SMART; SM00214; VMC; 1.
DR PROSITE, PS01208; VWF_C1; UNKNOWN 1.
DR PROSITE, PS0184; VWF_C2; 1.
KW Collagen.
SQ SEQUENCE 1464 AA; 138971 MW; 2B38BC27AF21590B CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1464;
Best Local Similarity 80.0%; Pred. No. 54;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|:|||||
Db 1169 GERGSEGSPPG 1178

RESULT 45
CO3A1 HUMAN
ID CO3A1_HUMAN STANDARD; PRT; 1466 AA.
AC P02461; Q15112; Q16403; Q6LDB3; Q6LDB3; Q7K256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN Name=COL3A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN NUCLEOTIDE SEQUENCE.
RC TISSUE=Skin fibroblast;
EX MEDLINE=93350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
chain of human type III procollagen. Differences in protein structure
from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516(1989).
RN NUCLEOTIDE SEQUENCE OF 149-1225.
RP MEDLINE=93386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1 (III)
collagen.";
RL Nucleic Acids Res. 17:6742-6742(1989).
RN NUCLEOTIDE SEQUENCE OF 168-398.
RP MEDLINE=77134724; PubMed=557335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
bromide peptides from the amino-terminal segment of type III collagen
of human liver.";
RL Biochemistry 16:1158-1164(1977).
RN NUCLEOTIDE SEQUENCE OF 186-194.
RP SEQUENCE REVISION.
RA Seyer J.M.;
RL Submitted (DEC-1977) to the PIR data bank.
RN NUCLEOTIDE SEQUENCE OF 186-194.
RX PubMed=8317500;
RA Milewicz D.M., Witz A.M., Smith A.C., Manchester D.K., Waldstein G.,
RA Byers P.H.;
RT "Parental somatic and germ-line mosaicism for a multixon deletion
with unusual endpoints in a type III collagen (COL3A1) allele produces
Ehlers-Danlos syndrome type IV in the heterozygous offspring.";
RT Am. J. Hum. Genet. 53:62-70(1993).
RN NUCLEOTIDE SEQUENCE OF 302-423.
RX PubMed=7487954;
RA Chido A.A., Sillence D.O., Cole W.G., Bateman J.F.;
RT "Abnormal type III collagen produced by an exon-17-skipping mutation
of the COL3A1 gene in Ehlers-Danlos syndrome type IV is not

```

RT incorporated into the extracellular matrix.";
RL Biochem. J. 311:939-943(1995).
[7]
RN
RP PROTEIN SEQUENCE OF 399-727.
RX MEDLINE=79000343; PubMed=687591;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBr peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411(1978).
[8]
RN
RP NUCLEOTIDE SEQUENCE OF 537-605.
RX PubMed=1672129;
RA Lee B., Vitale E., Superti-Furga A., Steinmann B., Ramirez F.;
RT "G to T transversion at position +5 of a splice donor site causes
RT skipping of the preceding exon in the type III procollagen transcripts
RT of a patient with Ehlers-Danlos syndrome type IV.";
RL J. Biol. Chem. 266:5256-5259(1991).
[9]
RN
RP PROTEIN SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1 (III)-
RT CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
[10]
RN
RP NUCLEOTIDE SEQUENCE OF 861-1015.
RX PubMed=2145268;
RA Cole W.G., Chiodo A.A., Lamande S.R., Janeczko R., Ramirez F.,
Dahl H.H., Chan D., Bateman J.F.;
RT "A base substitution at a splice site in the COL3A1 gene causes exon
RT skipping and generates abnormal type III procollagen in a patient with
RT Ehlers-Danlos syndrome type IV.";
RL J. Biol. Chem. 265:17070-17077(1990).
[11]
RN
RP NUCLEOTIDE SEQUENCE OF 950-1466.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalglish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337(1988).
[12]
RN
RP SEQUENCE REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalglish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
[13]
RN
RP NUCLEOTIDE SEQUENCE OF 1065-1466.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
[14]
RN
RP PROTEIN SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1(III)-
RT CB9 from type III collagen of human liver.";
RL Biochemistry 20:2621-2627(1981).
[15]
RN
RP NUCLEOTIDE SEQUENCE OF 1176-1466.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1(III)
RT collagen. Partial characterization of the 3' end region of the gene.";
RL J. Biol. Chem. 260:4357-4363(1985).
[16]
RN
RP NUCLEOTIDE SEQUENCE OF 1161-1200.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalglish R., Klueve-Beckerman B., Rennard S.I.,
Tolstoshev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
with the type I collagen genes during fibroblast growth.";
RL Biochemistry 25:1408-1413(1986).
[17]
RN
RP NUCLEOTIDE SEQUENCE OF 1165-1196.
RX MEDLINE=85216505; PubMed=3858826;
RA Emanuel B.S., Cannizzaro L.A., Seyer J.M., Myers J.C.;
RT "Human alpha 1(III) and alpha 2(V) procollagen genes are located on
RT the long arm of chromosome 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3385-3389(1985).
[18]
RN
RP NUCLEOTIDE SEQUENCE OF 1-170.
RX TISSUE=Placenta;
RA MEDLINE=88303360; PubMed=3405773;
RA Toman D., Ricca G., de Crombrughe B.;
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region of
RT human pro alpha 1(III) collagen.";
RL Nucleic Acids Res. 16:7201-7201(1988).
[19]
RN
RP NUCLEOTIDE SEQUENCE OF 1-176.
RX MEDLINE=89378752; PubMed=2777083; DOI=10.1016/0378-1119(89)90228-X;
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
RT procollagen gene (COL3A1).";
RL Gene 78:255-265(1989).
[20]
RN
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RX DOI=10.1002/(SICI)1098-1004(1997)9:4<300::AID-HUMU2>3.3.CO;2-8;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
[21]
RN
RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
RX MEDLINE=93293988; PubMed=8514866;
RA Tromp G., Wu Y., Prockop D.J., Madhathari S.L., Kleinert C.,
Barley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
Cole C.W., Jaakkola P., Ryyanen M., Pearce W.H., Yao J.S.F.,
Majamaa K., Smellens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
RT in the triple-helical domain of type III procollagen are an infrequent
RT cause of aortic aneurysms.";
RL J. Clin. Invest. 91:2539-2545(1993).
[22]
RN
RP VARIANT THR-698.
RX MEDLINE=91045136; PubMed=2235526;
RA Zafarullah K., Kleinert C., Tromp G., Kuivaniemi H., Kontusaari S.,
Wu Y., Ganguly A., Prockop D.J.;
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
RL Nucleic Acids Res. 18:6180-6180(1990).
[23]
RN
RP VARIANT AORTIC ANEURYSM ARG-786.
RX MEDLINE=91056145; PubMed=2243125;
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
RT "A mutation in the gene for type III procollagen (COL3A1) in a family
RT with aortic aneurysms.";
RL J. Clin. Invest. 86:1465-1473(1990).
[24]
RN
RP VARIANT EDS-IV ARG-828.
RX MEDLINE=94016385; PubMed=8411057;
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
RT "The substitution of glycine 661 by arginine in type III collagen
RT produces mutant molecules with different thermal stabilities and
RT causes Ehlers-Danlos syndrome type IV.";
RL J. Med. Genet. 30:690-693(1993).
[25]
RN
RP VARIANT EDS-IV SER-957.
RX MEDLINE=89109135; PubMed=2492273;
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;
Query Match 77.4%; Score 48; DB 1; Length 1466;

Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEXGAGSPG 10
Db 1170 GERGSESPG 1179

RESULT 46
Q53591 HUMAN
ID Q53591 HUMAN PRELIMINARY; PRT; 1466 AA.
AC Q53591;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein COL3A1.
GN Name=COL3A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Haakenson W., Pearson C.;
RT "The sequence of Homo sapiens BAC clone RP11-655A7.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.H.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Wilson R.K.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC066694; AAY24164.1; -; Genomic DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 1466 AA; 138564 MW; B904B4E05E17D339 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1466;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEXGAGSPG 10
Db 1170 GERGSESPG 1179

RESULT 47
Q541P8 HUMAN
ID Q541P8 HUMAN PRELIMINARY; PRT; 1466 AA.
AC Q541P8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Type III procollagen alpha 1 chain.
GN Name=COL3A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Haakenson W., Pearson C.;
RT "The sequence of Homo sapiens BAC clone RP11-655A7.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.H.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Wilson R.K.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC066694; AAY24164.1; -; Genomic DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 1466 AA; 138564 MW; B904B4E05E17D339 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1466;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEXGAGSPG 10
Db 1170 GERGSESPG 1179

RESULT 48
Q5DTG2 MOUSE
ID Q5DTG2 MOUSE PRELIMINARY; PRT; 1467 AA.
AC Q5DTG2;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE MKIAA4231 protein (Fragment).
GN Name=Col3a1; Synonyms=mkIAA4231;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA TISSUE=Fetal brain;
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene. The Complete Nucleotide Sequences of Mouse KIAA-homologous cDNAs Identified by Screening of Terminal sequences of cDNA Clones Randomly Sampled from Size-Fractionated Libraries.";
RT Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK220558; BAD90325.1; -; mRNA.
DR MGI; MGI:88453; Col3a1.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Cig_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 17.
DR ProDom; PD000007; Cig_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C; 1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1467 AA; 139335 MW; 73274A578803DCF8 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1467;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEXGAGSPG 10
Db 1170 GERGSESPG 1179

RESULT 49
Q541P8 HUMAN
ID Q541P8 HUMAN PRELIMINARY; PRT; 1466 AA.
AC Q541P8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Type III procollagen alpha 1 chain.
GN Name=COL3A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Haakenson W., Pearson C.;
RT "The sequence of Homo sapiens BAC clone RP11-655A7.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.H.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Wilson R.K.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC066694; AAY24164.1; -; Genomic DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 1466 AA; 138564 MW; B904B4E05E17D339 CRC64;

Query Match 77.4%; Score 48; DB 2; Length 1467;
Best Local Similarity 80.0%; Pred. No. 55;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GEXGAGSPG 10
Db 1170 GERGSESPG 1179

RESULT 50
Q541P8 HUMAN
ID Q541P8 HUMAN PRELIMINARY; PRT; 1466 AA.
AC Q541P8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Type III procollagen alpha 1 chain.
GN Name=COL3A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Haakenson W., Pearson C.;
RT "The sequence of Homo sapiens BAC clone RP11-655A7.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.H.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RA Waterston R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Wilson R.K.;
RL Submitted (APR-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC066694; AAY24164.1; -; Genomic DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 1466 AA; 138564 MW; B904B4E05E17D339 CRC64;

```

QY 1 GEKGAGSPG 10
DB 1172 GERSESGSPG 1181

RESULT 49
Q95JD7 MACMU PRELIMINARY; PRT; 243 AA.
AC Q95JD7 MACMU PRELIMINARY; PRT; 243 AA.
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adiponectin.
GN Name=APM1;
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
OC Cercopithecoidea; Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Adipose.
RX MEDLINE=2132234; PubMed=11334417;
RA Hotta K., Funahashi T., Bodkin N.L., Ortmeier H.K., Arita Y.,
Hansen B.C., Matsuzawa Y.;
RT "Circulating concentrations of the adipocyte protein adiponectin are
decreased in parallel with reduced insulin sensitivity during the
progression to type 2 diabetes in rhesus monkeys.";
RL Diabetes 50:1126-1133(2001).
DR EMBL; AF404407; AAK92202.1; -, mRNA.
DR HSP; Q60994; LC28.
DR SMK; Q95JD7; 107-243.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR008161; Clq helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF00386; Clq; 1.
DR PRINTS; PR00007; Collagen; 1.
DR PRODOM; PD000007; Clq helix; 1.
DR SMART; SM00110; Clq; 1.
KW Collagen.
SQ SEQUENCE 243 AA; 26264 MW; 49A45DAP2B4613FD CRC64;

Query Match 75.8%; Score 47; DB 2; Length 243;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSPG 12
DB 62 GEKGAGSPG 73

RESULT 50
ADIPO HUMAN STANDARD; PRT; 244 AA.
AC Q15848;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Adiponectin precursor (Adipocyte, Clq and collagen domain containing
protein) (30 kDa adipocyte complement-related protein) (ACRP30)
DE (Adipose most abundant gene transcript 1) (apM-1) (Gelatin-binding
protein).
GN Name=ADIPQ; Synonyms=ACDC, ACRP30, APM1, GBP28;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

```

RC TISSUE=Adipose tissue;
RX MEDLINE=96224171; PubMed=8619847; DOI=10.1006/bbrc.1996.0587;
RA Maeda K., Okubo K., Shimomura I., Funahashi T., Matsuzawa Y.,
Matsubara K.;
RT "cDNA cloning and expression of a novel adipose specific collagen-like
factor, apM1 (Adipose Most abundant Gene transcript 1).";
RL Biochem. Biophys. Res. Commun. 221:286-289(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99196984; PubMed=10095105; DOI=10.1016/S0378-1119(99)00041-4;
RA Saito K., Tobe T., Minoshima S., Asakawa S., Sumiya J., Yoda M.,
Nakano Y., Shimizu N., Tomita M.;
RT "Organization of the gene for gelatin-binding protein (GBP28).";
RL Gene 229:67-73(1999).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99333693; PubMed=10403784; DOI=10.1006/bbrc.1999.0865;
RA Schaeffler A., Orso E., Palitzsch K.D., Buechler C., Drobnik W.,
Fuerst A., Schoelmerich J., Schmitz G.;
RT "The human apM-1, an adipocyte-specific gene linked to the family of
TNF's and to genes expressed in activated T cells, is mapped to
chromosome 1q21.3-q23, a susceptibility locus identified for familial
combined hyperlipidemia (FCH).";
RL Biochem. Biophys. Res. Commun. 260:416-425(1999).
RN [4]
RP PROTEIN SEQUENCE OF 19-33.
RX PubMed=15340161; DOI=10.1110/ps.04682504;
RA Zhang Z., Henzel W.J.;
RT "Signal peptide prediction based on analysis of experimentally
verified cleavage sites.";
RL Protein Sci. 13:2819-2824(2004).
RN [5]
RP PROTEIN SEQUENCE OF N-TERMINUS, AND PARTIAL PROTEIN SEQUENCE.
RX PubMed=8947845;
RA Nakano Y., Tobe T., Choi-Miura N.H., Mazda T., Tomita M.;
RT "Isolation and characterization of GBP28, a novel gelatin-binding
protein purified from human plasma.";
RL J. Biochem. 120:803-812(1996).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=20417747; PubMed=10961870;
RA Yokota T., Oritani K., Takahashi I., Ishikawa J., Matsuyama A.,
Ouchi N., Kihara S., Funahashi T., Tenner A.J., Tomiyama Y.,
Matsuzawa Y.;
RT "Adiponectin, a new member of the family of soluble defense collagens,
negatively regulates the growth of myelomonocytic progenitors and the
functions of macrophages.";
RL Blood 96:1723-1732(2000).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE=20440368; PubMed=10982546;
RA Ouchi N., Kihara S., Arita Y., Okamoto Y., Maeda K., Kuriyama H.,
Hotta K., Nishida M., Takahashi M., Muraguchi M., Ohmoto Y.,
Nakamura T., Yamashita S., Funahashi T., Matsuzawa Y.;
RT "Adiponectin, an adipocyte-derived plasma protein, inhibits
endothelial NF-kappaB signaling through a CAMP-dependent pathway.";
RL Circulation 102:1296-1301(2000).
RN [8]
RP FUNCTION.
RX MEDLINE=21372498; PubMed=11479627; DOI=10.1038/90984;
RA Yamauchi T., Kamon J., Waki H., Terauchi Y., Kubota N., Hara K.,
Mori Y., Ide T., Murakami K., Tsuboyama-Kasaoka N., Ezaki O.,
Akanuma Y., Gavrilova O., Vinson C., Reitman M.L., Kagechika H.,
Shudo K., Yoda M., Nakano Y., Tobe K., Nagai R., Kimura S., Tomita M.,
Froguel P., Kadowaki T.;
RT "The fat-derived hormone adiponectin reverses insulin resistance
associated with both lipodystrophy and obesity.";
RL Nat. Med. 7:941-946(2001).
RN [9]
RP VARIANT ADIPONECTIN DEFICIENCY CYS-112.
RX MEDLINE=20378830; PubMed=10918532; DOI=10.1038/sj.ijo.0801244;
RA Takahashi M., Arita Y., Yamagata K., Matsukawa Y., Okutomi K.,
Horie M., Shimomura I., Hotta K., Kuriyama H., Kihara S., Nakamura T.,

```

RA Yamashita S., Funahashi T., Matsuzawa Y.;
 RT "Genomic structure and mutations in adipose-specific gene,
 RL Int. J. Obes. Relat. Metab. Disord. 24:861-868(2000).
 RN [10]
 RP VARIANTS ARG-84; MET-117; THR-164; SER-221 AND PRO-241.
 RX MEDLINE=21671103; PubMed=11812766;
 RA Hara K., Boutin P., Mori Y., Tobe K., Dina C., Yasuda K., Yamauchi T.,
 RA Otake S., Okada T., Eto K., Kadowaki H., Hagura R., Akanuma Y.,
 RA Yazaki Y., Nagai R., Taniyama M., Matsubara K., Yoda M., Nakano Y.,
 RA Kimura S., Tomita M., Kimura S., Ito C., Froguel P., Kadowaki T.;
 RT "Genetic variation in the gene encoding adiponectin is associated with
 an increased risk of type 2 diabetes in the Japanese population.";
 RL Diabetes 51:536-540(2002).
 CC -!- FUNCTION: Important negative regulator in hematopoiesis and immune
 CC systems; may be involved in ending inflammatory responses through
 CC its inhibitory functions. Inhibits endothelial NF-kappa-B
 CC signaling through a CAMP-dependent pathway. Inhibits TNF-alpha-
 CC induced expression of endothelial adhesion molecules. Involved in
 CC the control of fat metabolism and insulin sensitivity.
 CC -!- SUBUNIT: Homooligomer (Potential).
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: Synthesized exclusively by adipocytes and
 CC secreted into plasma.
 CC -!- DISEASE: Defects in ADIPOQ are the cause of adiponectin deficiency
 CC [MIM:605441]. The result is a very low concentration of plasma
 CC adiponectin. Decreased adiponectin plasma levels are associated
 CC with obesity insulin resistance, and diabetes type 2.
 CC -!- PHARMACEUTICAL: Adiponectin might be used in the treatment of
 CC diabetes type 2 and insulin resistance.
 CC -!- SIMILARITY: Contains 1 C1q domain.
 CC -!- SIMILARITY: Contains 1 collagen-like domain.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use as long as its content is in no way modified and this statement is not
 CC removed.
 DR EMBL; DA5371; BAA08227.1; -; mRNA.
 DR EMBL; AB012165; BAA86716.1; -; Genomic DNA.
 DR EMBL; AJ131460; CAB52413.1; -; Genomic DNA.
 DR EMBL; AJ131461; CAB52413.1; JOINED; Genomic DNA.
 DR PIR; JC4708; JC4708.
 DR HSP; O60954; 1C28.
 DR SRR; Q15848; 108-244.
 DR Ensembl; ENSG00000181092; Homo sapiens.
 DR HGNC; HGNC:13633; ADIPOQ.
 DR MIM; 605441; -
 DR GO; GO:0005576; C:extracellular region; NAS.
 DR GO; GO:0008091; P:generation of precursor metabolites and energy; TAS.
 DR InterPro; IPR001073; C1q.
 DR InterPro; IPR008161; C1q helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF00386; C1q; 1.
 DR Pfam; PF01391; Collagen; 1.
 DR PRINTS; PR00007; COMPLEXTQC1Q.
 DR ProDom; PD000007; C1q_helix; 2.
 DR PROSITE; PS00871; C1Q; 1.
 DR PROSITE; PS00871; C1Q; 1.
 KW Collagen; Diabetes mellitus; Direct protein sequencing;
 KW Disease mutation; Glycoprotein; Hormone; Hydroxylation; Obesity;
 KW Plasma; Polymorphism; Repeat; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 244 Adiponectin.
 FT DOMAIN 42 107 Collagen-like.
 FT DOMAIN 108 244 C1q.
 FT MOD_RES 44 44 4-hydroxyproline (By similarity).
 FT MOD_RES 47 47 4-hydroxyproline (By similarity).
 FT MOD_RES 53 53 4-hydroxyproline (By similarity).
 FT MOD_RES 62 62 4-hydroxyproline (By similarity).
 FT MOD_RES 71 71 4-hydroxyproline (By similarity).
 FT MOD_RES 76 76 4-hydroxyproline (By similarity).
 FT MOD_RES 86 86 4-hydroxyproline (By similarity).

FT MOD_RES 95 95 4-hydroxyproline.
 FT MOD_RES 104 104 4-hydroxyproline (By similarity).
 FT CARBOHYD 230 230 N-linked (GlcNAc...) (Potential).
 FT DISULFID 36 36 Interchain (By similarity).
 FT VARIANT 84 84 G -> R.
 FT FTID=VAR_013273.
 FT R -> C (in adiponectin deficiency).
 FT FTID=VAR_013274.
 FT V -> M.
 FT FTID=VAR_013275.
 FT I -> T.
 FT FTID=VAR_013276.
 FT R -> S.
 FT FTID=VAR_013277.
 FT H -> P.
 FT FTID=VAR_013278.
 FT SEQUENCE 244 AA; 26414 MW; 64D8C6C1204B1018 CRC64;
 SQ
 Query Match 75.8%; Score 47; DB 1; Length 244;
 Best Local Similarity 66.7%; Pred. No. 12;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GEKAGEGSPGLL 12
 DB 63 GEKAGEGDPGLI 74
 ||||| : |||||
 RESULT 51
 Q58EX9 HUMAN
 ID Q58EX9_HUMAN PRELIMINARY; PRT; 244 AA.
 AC Q58EX9;
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE ADIPOQ protein (Adiponectin).
 GN Name=ADIPOQ;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones, and Peripheral Nervous System;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Peripheral Nervous System;
 RG NIH MGC Project;
 RN Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=PCR rescued clones;


```
RG NIH MGC Project;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054496; AAH54496.1; -; mRNA.
DR EMBL; BC096309; AAH96309.1; -; mRNA.
DR EMBL; BC096310; AAH96310.1; -; mRNA.
DR EMBL; BC096311; AAH96311.1; -; mRNA.
DR EMBL; BC096308; AAH96308.1; -; mRNA.
DR SMR; Q58EX9; 108-244.
DR Ensembl; ENSG00000181092; Homo sapiens.
DR HGNC; HGNC:13633; ADIPOQ.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR001073; Clg.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF00386; Clg; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLEMNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS00871; C1Q; 1.
KW COLLAGEN.
SQ SEQUENCE 244 AA; 26414 MW; 64D8C6C1204B1018 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGLL 12
Db 63 GEKGAGSPGLI 74

RESULT 52
Q5TMG9 ANOGA PRELIMINARY; PRT; 375 AA.
AC Q5TMG9;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE ENSANGP0000026445 (Fragment).
GN ORFNames=ENSANG0000016690;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008986; EAA00623.2; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 13.
DR ProDom; PD000007; Clg_helix; 2.
KW COLLAGEN.
FT NON TER 1
FT NON TER 802
FT NON TER 802
SQ SEQUENCE 802 AA; 73832 MW; 17C76A29EF52C679 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 802;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
Db 131 GEKGAGSPGI 141

RESULT 54
Q4SB07 TETNG PRELIMINARY; PRT; 856 AA.
AC Q4SB07;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAFI4677, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG00021178001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontoidea; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]

Query Match 75.8%; Score 47; DB 2; Length 375;
Best Local Similarity 72.7%; Pred. No. 19;

RG NIH MGC Project;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC054496; AAH54496.1; -; mRNA.
DR EMBL; BC096309; AAH96309.1; -; mRNA.
DR EMBL; BC096310; AAH96310.1; -; mRNA.
DR EMBL; BC096311; AAH96311.1; -; mRNA.
DR EMBL; BC096308; AAH96308.1; -; mRNA.
DR SMR; Q58EX9; 108-244.
DR Ensembl; ENSG00000181092; Homo sapiens.
DR HGNC; HGNC:13633; ADIPOQ.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR001073; Clg.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF00386; Clg; 1.
DR Pfam; PF01391; Collagen; 1.
DR PRINTS; PR00007; COMPLEMNTC1Q.
DR SMART; SM00110; C1Q; 1.
DR PROSITE; PS00871; C1Q; 1.
KW COLLAGEN.
SQ SEQUENCE 244 AA; 26414 MW; 64D8C6C1204B1018 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 12;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGLL 12
Db 63 GEKGAGSPGLI 74

RESULT 52
Q5TMG9 ANOGA PRELIMINARY; PRT; 375 AA.
AC Q5TMG9;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE ENSANGP0000026445 (Fragment).
GN ORFNames=ENSANG0000016690;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Culicidae;
OC Anophelinae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PEST;
RG The Anopheles gambiae Sequence Committee;
RT "Anopheles gambiae re-annotation.";
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAB01008986; EAA00623.2; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Clg_helix; 2.
KW COLLAGEN.
FT NON TER 1
FT NON TER 375
FT NON TER 375
SQ SEQUENCE 375 AA; 34271 MW; 00D5FBE48ED834EF CRC64;

Query Match 75.8%; Score 47; DB 2; Length 375;
Best Local Similarity 72.7%; Pred. No. 19;
```

Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C., Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.; "Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype."; Nature 431:946-957(2004).

[2]

NRPN NUCLEOTIDE SEQUENCE.

RRP Genoscope; Whitehead Institute Centre for Genome Research; Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.

CCC -!- CAUTION: The sequence shown here is derived from an EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is preliminary data.

CC EMBL; CAAG01011805; CAP94059.1; -; Genomic_DNA.

CC InterPro; IPR008161; C1g_helix.

CDR InterPro; IPR008160; Collagen.

CDR InterPro; IPR001442; Procollagen4_C.

CDR Pfam; PF01413; C4; 2.

CDR Pfam; PF01391; Collagen; 10.

CDR ProDom; PD000007; C1g_helix; 4.

CDR ProbDom; PD003923; Procollagen4_C; 2.

CDR SMART; SM00111; C4; 2.

KW Collagen.

FT NON_TER 1026 1026

SQ SEQUENCE 1026 AA; 103855 MW; 1D61B3PF1779C367 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 1026;
Best Local Similarity 72.7%; Pred. NO. 56;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSGPGL 11
| : : : : :
DB 451 GRRGAGSGPGL 461

RESULT 56

CAFF_RIFPA STANDARD; PRT; 1027 AA.

ID CAFF_RIFPA STANDARD; PRT; 1027 AA.

AC P30754;

DT 01-JUL-1993 (Rel. 26, Created)

DT 25-OCT-2004 (Rel. 45, Last sequence update)

DT 10-MAY-2005 (Rel. 47, Last annotation update)

DE Fibril-forming collagen alpha chain (Fragment).

OS Riftia pachyptila (tube worm).

OC Eukaryota; Metazoa; Pogonophora; Vestimentifera; Axonobranchia;

OC Riftiida; Riftiidae; Riftia.

OX NCBI_TaxID=6426;

RN [1]

RP PROTEIN SEQUENCE.

RX MEDLINE=93130909; PubMed=1483468;

RA Mann K., Gaill F., Timpl R.;

RT "Amino-acid sequence and cell-adhesion activity of a fibril-forming collagen from the tube worm Riftia pachyptila living at deep sea hydrothermal vents.";

RT Eur. J. Biochem. 210:839-847(1992).

RL [2]

RN PROTEIN SEQUENCE OF 8-45; 525-618 AND 810-882.

RC TISSUE=Cuticle;

RC MEDLINE=92015209; PubMed=1920405;

RA Gaill F., Wiedemann H., Mann K., Kuhn K., Timpl R., Engel J.;

RT "Molecular characterization of cuticle and interstitial collagens from worms collected at deep sea hydrothermal vents.";

RT J. Mol. Biol. 221:209-223(1991).

CC -!- FUNCTION: Fibril-forming collagen.

CC -!- SUBUNIT: Homotetramer.

CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC PIR; S28774; S28774.

DR InterPro; IPR008161; C1g_helix.

DR InterPro; IPR008160; Collagen.
DR Pfam; PF01391; Collagen; 16.
DR ProDom; PD000007; Clg_helix; 8.
KW Collagen; Direct protein sequencing; Extracellular matrix;
KW Glycoprotein; Repeat; Structural protein.
FT REGION 1 12 Nonhelical region (N-terminal).
FT REGION 13 1023 Triple-helical region.
FT REGION 1024 1027 Nonhelical region (C-terminal).
FT SITE 610 610 Imperfection in the GAA repeat.
FT MOD_RES 21 21 4-hydroxyproline (partial).
FT MOD_RES 24 24 4-hydroxyproline (partial).
FT MOD_RES 27 27 4-hydroxyproline.
FT MOD_RES 39 39 4-hydroxyproline.
FT MOD_RES 53 53 3-hydroxyproline (partial).
FT MOD_RES 54 54 4-hydroxyproline.
FT MOD_RES 72 72 4-hydroxyproline (partial).
FT MOD_RES 90 90 4-hydroxyproline.
FT MOD_RES 93 93 4-hydroxyproline.
FT MOD_RES 96 96 5-hydroxylysine (Probable).
FT MOD_RES 108 108 5-hydroxylysine (Probable).
FT MOD_RES 123 123 4-hydroxyproline (partial).
FT MOD_RES 128 128 4-hydroxyproline (partial).
FT MOD_RES 150 150 4-hydroxyproline.
FT MOD_RES 161 161 3-hydroxyproline (partial).
FT MOD_RES 162 162 4-hydroxyproline.
FT MOD_RES 164 164 4-hydroxyproline (partial).
FT MOD_RES 165 165 3-hydroxyproline.
FT MOD_RES 174 174 4-hydroxyproline.
FT MOD_RES 177 177 4-hydroxyproline.
FT MOD_RES 180 180 4-hydroxyproline.
FT MOD_RES 183 183 5-hydroxylysine.
FT MOD_RES 192 192 5-hydroxylysine (Probable).
FT MOD_RES 207 207 4-hydroxyproline.
FT MOD_RES 216 216 4-hydroxyproline.
FT MOD_RES 219 219 4-hydroxyproline.
FT MOD_RES 228 228 4-hydroxyproline.
FT MOD_RES 237 237 4-hydroxyproline.
FT MOD_RES 243 243 4-hydroxyproline (partial).
FT MOD_RES 249 249 4-hydroxyproline.
FT MOD_RES 255 255 4-hydroxyproline.
FT MOD_RES 261 261 5-hydroxylysine (Probable).
FT MOD_RES 273 273 4-hydroxyproline (partial).
FT MOD_RES 276 276 4-hydroxyproline (partial).
FT MOD_RES 279 279 5-hydroxylysine (Probable).
FT MOD_RES 285 285 4-hydroxyproline (partial).
FT MOD_RES 291 291 4-hydroxyproline (partial).
FT MOD_RES 303 303 4-hydroxyproline (partial).
FT MOD_RES 306 306 4-hydroxyproline.
FT MOD_RES 312 312 4-hydroxyproline.
FT MOD_RES 321 321 4-hydroxyproline.
FT MOD_RES 327 327 4-hydroxyproline.
FT MOD_RES 339 339 4-hydroxyproline.
FT MOD_RES 342 342 5-hydroxylysine.
FT MOD_RES 348 348 4-hydroxyproline (partial).
FT MOD_RES 351 351 5-hydroxylysine (partial).
FT MOD_RES 366 366 4-hydroxyproline.
FT MOD_RES 372 372 4-hydroxyproline.
FT MOD_RES 375 375 4-hydroxyproline.
FT MOD_RES 381 381 4-hydroxyproline (partial).
FT MOD_RES 387 387 4-hydroxyproline.
FT MOD_RES 416 416 3-hydroxyproline (partial).
FT MOD_RES 417 417 4-hydroxyproline.
FT MOD_RES 423 423 4-hydroxyproline.
FT MOD_RES 429 429 4-hydroxyproline.
FT MOD_RES 432 432 4-hydroxyproline.
FT MOD_RES 453 453 4-hydroxyproline.
FT MOD_RES 465 465 4-hydroxyproline.
FT MOD_RES 483 483 4-hydroxyproline.
FT MOD_RES 500 500 4-hydroxyproline (partial).
FT MOD_RES 503 503 4-hydroxyproline (partial).
FT MOD_RES 506 506 4-hydroxyproline (partial).
FT MOD_RES 513 513 4-hydroxyproline.
FT MOD_RES 525 525 4-hydroxyproline.
FT MOD_RES 533 533 4-hydroxyproline (partial).
FT MOD_RES 536 536 4-hydroxyproline (partial).
FT MOD_RES 540 540 4-hydroxyproline.
FT MOD_RES 546 546 5-hydroxylysine.
FT MOD_RES 551 551 3-hydroxyproline (partial).
FT MOD_RES 552 552 4-hydroxyproline.
FT MOD_RES 561 561 4-hydroxyproline.
FT MOD_RES 567 567 5-hydroxylysine.
FT MOD_RES 573 573 5-hydroxylysine (Probable).
FT MOD_RES 603 603 4-hydroxyproline.
FT MOD_RES 612 612 5-hydroxylysine (Probable).
FT MOD_RES 621 621 4-hydroxyproline (partial).
FT MOD_RES 627 627 4-hydroxyproline.
FT MOD_RES 645 645 4-hydroxyproline (partial).
FT MOD_RES 647 647 3-hydroxyproline (partial).
FT MOD_RES 648 648 4-hydroxyproline.
FT MOD_RES 657 657 5-hydroxylysine (Probable).
FT MOD_RES 663 663 4-hydroxyproline.
FT MOD_RES 708 708 4-hydroxyproline.
FT MOD_RES 711 711 4-hydroxyproline.
FT MOD_RES 714 714 4-hydroxyproline.
FT MOD_RES 717 717 4-hydroxyproline.
FT MOD_RES 723 723 4-hydroxyproline.
FT MOD_RES 738 738 5-hydroxylysine (Probable).
FT MOD_RES 744 744 4-hydroxyproline.
FT MOD_RES 759 759 4-hydroxyproline.
FT MOD_RES 765 765 5-hydroxylysine (Probable).
FT MOD_RES 773 773 3-hydroxyproline (partial).
FT MOD_RES 774 774 4-hydroxyproline.
FT MOD_RES 783 783 4-hydroxyproline.
FT MOD_RES 792 792 4-hydroxyproline.
FT MOD_RES 810 810 5-hydroxylysine (Probable).
FT MOD_RES 815 815 3-hydroxyproline (partial).
FT MOD_RES 816 816 4-hydroxyproline.
FT MOD_RES 843 843 4-hydroxyproline.
FT MOD_RES 849 849 4-hydroxyproline.
FT MOD_RES 855 855 4-hydroxyproline.
FT MOD_RES 861 861 4-hydroxyproline.
FT MOD_RES 867 867 4-hydroxyproline.
FT MOD_RES 888 888 4-hydroxyproline.
FT MOD_RES 894 894 4-hydroxyproline.
FT MOD_RES 903 903 4-hydroxyproline.
FT MOD_RES 915 915 4-hydroxyproline.
FT MOD_RES 927 927 5-hydroxylysine (Probable).
FT MOD_RES 933 933 5-hydroxylysine (partial).
FT MOD_RES 936 936 5-hydroxylysine (Probable).
FT MOD_RES 939 939 5-hydroxylysine.
FT MOD_RES 945 945 4-hydroxyproline (partial).
FT MOD_RES 954 954 4-hydroxyproline.
FT MOD_RES 963 963 4-hydroxyproline.
FT MOD_RES 966 966 4-hydroxyproline.
FT MOD_RES 984 984 4-hydroxyproline.
FT MOD_RES 990 990 4-hydroxyproline (partial).
FT MOD_RES 1010 1010 3-hydroxyproline (partial).
FT MOD_RES 1011 1011 4-hydroxyproline.
FT MOD_RES 1013 1013 3-hydroxyproline (partial).
FT MOD_RES 1014 1014 4-hydroxyproline (partial).
FT MOD_RES 1016 1016 3-hydroxyproline (partial).
FT MOD_RES 1017 1017 4-hydroxyproline.
FT MOD_RES 1019 1019 3-hydroxyproline (partial).
FT MOD_RES 1020 1020 4-hydroxyproline.
FT CARBOHYD 96 O-linked (Gal. . .) (Probable).
FT CARBOHYD 108 O-linked (Gal. . .) (Probable).
FT CARBOHYD 192 O-linked (Gal. . .) (Probable).
FT CARBOHYD 261 O-linked (Gal. . .) (Probable).
FT CARBOHYD 279 O-linked (Gal. . .) (Probable).
FT CARBOHYD 573 O-linked (Gal. . .) (Probable).
FT CARBOHYD 612 O-linked (Gal. . .) (Probable).
FT CARBOHYD 657 O-linked (Gal. . .) (Probable).
FT CARBOHYD 738 O-linked (Gal. . .) (Probable).
FT CARBOHYD 765 O-linked (Gal. . .) (Probable).
FT CARBOHYD 810 O-linked (Gal. . .) (Probable).
FT CARBOHYD 927 O-linked (Gal. . .) (Probable).

```

FT CARBOHYD 936 936 O-linked (Gal. . .) (Probable).
FT VARIANT 903 903 P -> A.
FT UNSURE 96 96
FT UNSURE 108 108
FT UNSURE 192 192
FT UNSURE 261 261
FT UNSURE 279 279
FT UNSURE 573 573
FT UNSURE 612 612
FT UNSURE 657 657
FT UNSURE 738 738
FT UNSURE 765 765
FT UNSURE 810 810
FT UNSURE 927 927
FT UNSURE 936 936
FT NON_TER 1 1
FT NON_TER 1027 1027

Query Match 75.8%; Score 47; DB 1; Length 1027;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSPGLL 12
Db 736 GVGKAGSPGLV 747

RESULT 57
Q35053 MOUSE PRELIMINARY; PRT; 1136 AA.
AC Q35053;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Collagen al XIX chain.
GN Name=Coll19a1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=BALB/c;
RX MEDLINE=97347522; PubMed=9202028; DOI=10.1074/jbc.272.27.17104;
RA Sumiyoguchi H., Inoguchi K., Khaleduzzaman M., Ninomiya Y., Yoshioka H.;
RT "Ubiquitous expression of the alpha1(XIX) collagen gene (Coll19a1)
during mouse embryogenesis becomes restricted to a few tissues in the
adult organism.";
RL J. Biol. Chem. 272:17104-17111 (1997).
DR EMBL; AB000636; BAA23578.1; -; mRNA.
DR Ensembl; ENSMUSG0000026141; Mus musculus.
DR MGI; MGI:1095415; Coll19a1.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. . .; IMP.
DR GO; GO:0007519; P:myogenesis; IMP.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 12.
DR ProDom; PD000007; C1g_helix; 5.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Extracellular matrix; Structural protein.
SQ SEQUENCE 1136 AA; 114354 MW; B211A0135572FA97 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 1136;
Best Local Similarity 75.0%; Pred. No. 62;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEKGAGSPGLL 12
Db 483 GEKGAGSPGLL 494

```

```

RESULT 58
Q4SK58 TETNG PRELIMINARY; PRT; 1333 AA.
ID Q4SK58;
AC Q4SK58;
DT 13-SEP-2005 (TREMBlrel. 31, Created)
DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)
DE Chromosome 2 SCAF14570, whole genome shotgun sequence.
DE (Fragment).
GN ORENAMES=GSTENG00016870001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouteau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anhouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biewnot C., Skalli Z., Catolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chapple C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
the early vertebrate proto-karyotype.";
RL Nature 431:946-957 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RG Genoscope; Whitehead
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; CAAB01014570; CAF98974.1; -; Genomic_DNA.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR010515; Endostatin.
DR InterPro; IPR003129; Laminin_G_TSP_N.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF06482; Endostatin; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Extracellular matrix.
FT NON_TER 1 1
FT NON_TER 1333 1333
FT NON_TER 1333 AA; 137051 MW; 1355C59C0A8BFD21 CRC64;
SQ SEQUENCE 1333 AA; 137051 MW; 1355C59C0A8BFD21 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 1333;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSPGL 11
Db 499 GQKGEQSPGL 509

RESULT 59
Q60444 CRIGR PRELIMINARY; PRT; 1549 AA.
ID Q60444;
AC Q60444;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Type VII collagen (Fragment).
DE Cricetus griseus (Chinese hamster).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;

```

OC Muridae; Cricetinae; Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Lung;
 RX MEDLINE=93271985; PubMed=8499916;
 RA Greenspan D.S.;
 RT "The carboxyl-terminal half of type VII collagen, including the non-
 RT collagenous NC-2 domain and intron/exon organization of the
 RT corresponding region of the COL7A1 gene.";
 RL Hum. Mol. Genet. 2:273-278(1993).
 CC -1- SIMILARITY: Contains 1 BPTI/Kunitz inhibitor domain.
 DR EMBL; L06863; AAA36968.1; -; mRNA.
 DR PIR; I48103; I48103.
 DR HSSP; P00981; 1DTK.
 DR GO; GO:0005737; Cytoplasm; IEA.
 DR GO; GO:0004867; P:serine-type endopeptidase inhibitor activity; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR002223; Prot_inh_Kunz-m.
 DR Pfam; PF01391; Collagen; 22.
 DR Pfam; PF00014; Kunitz_BPTI; 1.
 DR PRINTS; PR00759; BASICTPASE.
 DR ProDom; PD000007; Clg_helix; 12.
 DR ProDom; PD000222; Prot_inh_Kunz-m; 1.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE; PS00279; BPTI_KUNITZ_2; 1.
 KW Collagen.
 FT NON TER 1 1
 SQ SEQUENCE 1549 AA; 148117 MW; 5D646391E0C29292 CRC64;
 Query Match 75.8%; Score 47; DB 2; Length 1549;
 Best Local Similarity 80.0%; Pred. No. 85;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GEKGAEGSPG 10
 Db 242 GEKGVGNPG 251
 ||||| |||||
 RESULT 60
 CO4A5 HUMAN
 ID CO4A5 HUMAN STANDARD; PRT; 1685 AA.
 AC P29400; Q16006; Q16126; Q6LD84;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-MAY-2005 (Rel. 47, Last annotation update)
 DE Collagen alpha 5(IV) chain precursor.
 GN Name=COL4A5;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=94165049; PubMed=8120014;
 RA Zhou J., Leinonen A., Tryggvason K.;
 RT "Structure of the human type IV collagen COL4A5 gene.";
 RL J. Biol. Chem. 269:6608-6614(1994).
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 1-910, AND VARIANT AS CVS-521.
 RC TISSUE=Kidney;
 RX MEDLINE=92316923; PubMed=1352287;
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain
 RT and identification of a single-base mutation in exon 23 converting
 RT glycine 521 in the collagenous domain to cysteine in an Alport
 RT syndrome patient.";
 RL J. Biol. Chem. 267:12475-12481(1992).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 85-1685.
 RC TISSUE=Placenta;
 RX MEDLINE=90337990; PubMed=2380186;
 RA Pihlajaniemi T., Pohjola-Einoja E.R., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the
 RT carboxyl-terminal domain of a new type IV collagen chain, alpha
 RT 5(IV).";
 RL J. Biol. Chem. 265:13758-13766(1990).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 924-1685.
 RX MEDLINE=91169491; PubMed=2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of the 3' half of the human type IV collagen alpha 5
 RT gene that is affected in the Alport syndrome.";
 RL Genomics 9:1-9(1991).
 RN [5]
 RP NUCLEOTIDE SEQUENCE OF 914-1685.
 RX MEDLINE=90160375; PubMed=1689491;
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
 RA Tryggvason K.;
 RT "Identification of a distinct type IV collagen alpha chain with
 RT restricted kidney distribution and assignment of its gene to the locus
 RT of X chromosome-linked Alport syndrome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 RN [6]
 RP NUCLEOTIDE SEQUENCE OF 1442-1471.
 RX MEDLINE=90252791; PubMed=2339699;
 RA Myers J.C., Jones T.A., Pohjola-Einoja E.R., Kadri A.S., Goddard A.D.,
 RA Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
 RT to the region of the X chromosome containing the Alport syndrome
 RT locus.";
 RL Am. J. Hum. Genet. 46:1024-1033(1990).
 RN [7]
 RP NUCLEOTIDE SEQUENCE OF 1-20.
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J.,
 RA Marynen P.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP NUCLEOTIDE SEQUENCE OF 1258-1270 (ISOFORM 2).
 RX MEDLINE=94133540; PubMed=8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
 RA Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells:
 RT a complex mutation in the COL4A5 gene of an Alport patient deletes the
 RT NCI domain.";
 RL Kidney Int. 44:1316-1321(1993).
 RN [9]
 RP NUCLEOTIDE SEQUENCE OF 1589-1598 AND 1677-1685, AND VARIANTS AS
 RX 1597-TYR--THR-1685 DEL AND 1679-GLQ--THR-1685 DEL.
 RA Nakazato H., Hattori S., Ushijima T., Matsuura T., Koitabashi Y.,
 RA Takada T., Yoshioka K., Endo F., Matsuda I.;
 RT "Mutations in the COL4A5 gene in Alport syndrome: a possible mutation
 RT in primordial germ cells.";
 RL Kidney Int. 46:1307-1314(1994).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RX DOI=10.1002/(SICI)1098-1004(1997)9:6<477::AID-HUMU13.3.CO;2-H;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 RN [11]
 RP VARIANT AS SER-1564.
 RX MEDLINE=91169492; PubMed=1672282;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
 RA Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
 RT conserved cysteine to serine in Alport syndrome.";
 RL Genomics 9:10-18(1991).
 RN [12]
 RP VARIANT AS ARG-325.
 RX MEDLINE=92303559; PubMed=1376965;

RA Knobelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P.,
RA Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV)
RT chain associated with X-linked Alport syndrome: characterization of
RT the mutation by direct sequencing of PCR-amplified lymphoblast cDNA
RT fragments."; Am. J. Hum. Genet. 51:135-142 (1992).
RL [13]
RN VARIANT AS GLU-325.
RP MEDLINE=93244772; PubMed=1363780;
RX Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L.,
RA Rizzoni G.F., de Marchi M.;
RA "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome."; Hum. Mol. Genet. 1:127-129 (1992).
RL [14]
RN VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of patients
RT with Alport syndrome."; Genomics 17:485-489 (1993).
RL [15]
RN VARIANTS AS GLU-400; VAL-406; VAL-638; ARG-653; ARG-796;
RP ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients."; Hum. Mutat. 5:197-204 (1995).
RL [16]
RN VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States."; Am. J. Hum. Genet. 58:1157-1165 (1996).
RL [17]
RN VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Bruttini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolarì F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene."; Am. J. Hum. Genet. 58:1192-1204 (1996).
RL [18]
RN VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knobelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome."; Am. J. Hum. Genet. 59:1221-1232 (1996).
RL [19]
RN VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA DOI=10.1002/(SICI)1098-1004(1996)7:2<149: AID-HUMU9>3.3.CO;2-A;
RA Tverskaya S., Bobrynya V., Tsalykova F., Ignatova M.,
RA Krasnopol'skaya X., Evgrafov O.;
RT "Substitution of Al98D in noncollagen domain of alpha 5(IV) collagen chain
RT associated with adult-onset X-linked Alport syndrome."; Hum. Mutat. 7:149-150 (1996).
RL [20]
RN VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741; DOI=10.1007/s004390050429;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;

RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q."; Hum. Genet. 99:681-684 (1997).
RL [21]
RN VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome."; Hum. Mutat. Suppl. 1:S106-S109 (1998).
RL [22]
RN VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY--GLU-807 DEL; ARG-869; CYS-941; SER-1030; SER-1066; ASP-1143;
RX ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RA MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykarsdottir U., Neumann H.,
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;
RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing."; J. Am. Soc. Nephrol. 9:2291-2301 (1998).
RL [23]
RN Query Match 75.8%; Score 47; DB 1; Length 1685;
RP Best Local Similarity 72.7%; Pred. No. 93;
RX Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
RY 1 GEXCAEGSPGL 11
RL 938 GEXGSGEPGL 948

RESULT 61
Q9N0B7_HUMAN
ID Q9N0B7_HUMAN PRELIMINARY; PRT; 1685 AA.
AC Q9N0B7; Q72700;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Collagen, type IV, alpha 5 (Alport syndrome).
GN Name=COL4A5; ORFNames=RP6-24A23.5-001;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Cobley V.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Bird C.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RX EMBL; AL035425; CAB30289.2; -; Genomic DNA.
DR EMBL; AL034369; CAA22287.2; -; Genomic DNA.
DR EMBL; AL031622; CAI43038.1; -; Genomic DNA.
DR EMBL; AL034369; CAB30289.2; JOINED; Genomic DNA.
DR EMBL; AL031622; CAB30289.2; JOINED; Genomic DNA.
DR EMBL; AL031622; CAA22267.2; JOINED; Genomic DNA.
DR EMBL; AL035425; CAA22267.2; JOINED; Genomic DNA.
DR EMBL; AL034369; CAI43038.1; JOINED; Genomic DNA.
DR EMBL; AL035425; CAI43038.1; JOINED; Genomic DNA.
DR SMR; Q9N0B7; 1458-1685.
DR Ensembl; ENSG00000188153; Homo sapiens.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Cig helix.
DR InterPro; IPR008160; Collagen.

```
DR InterPro: IPR001442; Procollagn4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 22.
DR ProDom: PD000007; Clg_helix; 3.
DR ProDom: PD0031923; Procollagn4_C; 2.
DR SMART: SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1685 AA; 161044 MW; 4450A6762F12A626 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 1685;
Best Local Similarity 72.7%; Pred. No. 93;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 11
|||||:||||
Db 938 GEGKSGKEPGL 948

RESULT 62
Q16299 HUMAN PRELIMINARY; PRT; 53 AA.
AC Q16299;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Type II collagen alpha 1 chain protein (Fragment).
GN Name=type II collagen alpha 1 chain;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95213626; PubMed=7699294;
RA Nakamura M., Nakamura M., Okazaki K., Masumi S.;
RT "Analysis of a Hind III site polymorphism in the type II collagen
RT gene: it's location and frequencies in the Japanese population.";
RL Nippon Seikeigeka Gakka Zasshi 69:11-16(1995).
DR EMBL: S76829; AAD14239.1; -; Genomic DNA.
DR GO: 0005515; F:protein binding; IEA.
DR GO: 00007155; P:cell adhesion; IEA.
DR InterPro: IPR008161; Clg_helix.
DR ProDom: PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER 53
SQ SEQUENCE 53 AA; 5323 MW; B7B11FB64D232B69 CRC64;

Query Match 74.2%; Score 46; DB 2; Length 53;
Best Local Similarity 80.0%; Pred. No. 3.7;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 10
|||||:||||
Db 40 GEGKPEGAPG 49

RESULT 63
Q866A5_PIG PRELIMINARY; PRT; 142 AA.
AC Q866A5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative collagen type XI alpha 1 (Fragment).
GN Name=COL11A1;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.

DR InterPro: IPR001442; Procollagn4_C.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 22.
DR ProDom: PD000007; Clg_helix; 3.
DR ProDom: PD0031923; Procollagn4_C; 2.
DR SMART: SM00111; C4; 2.
KW Collagen.
SQ SEQUENCE 1685 AA; 161044 MW; 4450A6762F12A626 CRC64;

Query Match 75.8%; Score 47; DB 2; Length 1685;
Best Local Similarity 72.7%; Pred. No. 93;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 11
|||||:||||
Db 938 GEGKSGKEPGL 948

RESULT 62
Q16299 HUMAN PRELIMINARY; PRT; 53 AA.
AC Q16299;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Type II collagen alpha 1 chain protein (Fragment).
GN Name=type II collagen alpha 1 chain;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95213626; PubMed=7699294;
RA Nakamura M., Nakamura M., Okazaki K., Masumi S.;
RT "Analysis of a Hind III site polymorphism in the type II collagen
RT gene: it's location and frequencies in the Japanese population.";
RL Nippon Seikeigeka Gakka Zasshi 69:11-16(1995).
DR EMBL: S76829; AAD14239.1; -; Genomic DNA.
DR GO: 0005515; F:protein binding; IEA.
DR GO: 00007155; P:cell adhesion; IEA.
DR InterPro: IPR008161; Clg_helix.
DR ProDom: PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER 53
SQ SEQUENCE 53 AA; 5323 MW; B7B11FB64D232B69 CRC64;

Query Match 74.2%; Score 46; DB 2; Length 53;
Best Local Similarity 80.0%; Pred. No. 3.7;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 10
|||||:||||
Db 40 GEGKPEGAPG 49

RESULT 63
Q866A5_PIG PRELIMINARY; PRT; 142 AA.
AC Q866A5;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative collagen type XI alpha 1 (Fragment).
GN Name=COL11A1;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
```

```
RC TISSUE=Cartilage;
RA Petersen J.P., Amling M., Meenen N.M., Haberland M.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ536287; CAD60250.1; -; mRNA.
DR GO: 0005737; C:cytoplasm; IEA.
DR GO: 0005515; F:protein binding; IEA.
DR GO: 0007155; P:cell adhesion; IEA.
DR GO: 0006817; P:phosphate transport; IEA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR Pfam: PF01391; Collagen; 2.
DR ProDom: PD000007; Clg_helix; 1.
KW Collagen.
FT NON_TER 142
SQ SEQUENCE 142 AA; 13498 MW; 4C9BFC87EDC7B77D CRC64;

Query Match 74.2%; Score 46; DB 2; Length 142;
Best Local Similarity 72.7%; Pred. No. 10;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 11
|||||:||||
Db 125 GEGKPGKPGGL 135

RESULT 64
Q9TT85_PIG PRELIMINARY; PRT; 167 AA.
AC Q9TT85;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Type II collagen alpha 1 (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
OC Sus.
OX NCBI_TaxID=9823;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX Wang J.Y., Baer A.E., Kraus V.B., Setton L.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF201724; AAF23131.1; -; mRNA.
DR GO: 0005737; C:cytoplasm; IEA.
DR GO: 0005515; F:protein binding; IEA.
DR GO: 0007155; P:cell adhesion; IEA.
DR GO: 0006817; P:phosphate transport; IEA.
DR InterPro: IPR008161; Clg_helix.
DR InterPro: IPR008160; Collagen.
DR Pfam: PF01391; Collagen; 2.
DR ProDom: PD000007; Clg_helix; 2.
KW Collagen.
FT NON_TER 167
SQ SEQUENCE 167 AA; 15250 MW; 5BC11178626AED93 CRC64;

Query Match 74.2%; Score 46; DB 2; Length 167;
Best Local Similarity 80.0%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 10
|||||:||||
Db 146 GEGKPEGAPG 155

RESULT 65
Q92416_CAVPO PRELIMINARY; PRT; 184 AA.
AC Q92416;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
```


Alpha-1 type II collagen (Fragment).
 DE Cavia porcellus (Guinea pig).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
 OC Hystriognathini; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX MEDLINE=21841527; PubMed=11852233; DOI=10.1016/S0945-053X(01)00193-7;
 RA Clark A.G., Rohrbach A.L., Otterness I., Kraus V.B.;
 RT "The effects of ascorbic acid on cartilage metabolism in guinea pig
 RT articular cartilage explants";
 RL Matrix Biol. 21:175-184(2002).
 DR EMBL; AF299351; AAK95495.1; -; mRNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR004829; Caurface_antigen.
 DR Pfam; PF01391; Collagen; 3.
 DR ProDom; PD000007; C1g_helix; 2.
 DR ProDom; PD153432; Caurface_antigen; 1.
 KW Collagen.
 FT NON TER 184 184
 FT NON TER 184 184
 SQ SEQUENCE 184 AA; 16706 MW; 1CA9802AD35DECA6 CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 184;
 Best Local Similarity 80.0%; Pred. No. 14;
 Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;
 OY 1 GEGKAGSPG 10
 DB 150 GEGKEGAPG 159
 RESULT 66
 O9GLK9 RABBIT PRELIMINARY; PRT; 187 AA.
 ID Q9GLK9
 AC Q9GLK9
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Alpha 1 type X collagen (Fragment).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Lagomorpha; Leporidae;
 OC Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RX Bluteau G., Mathieu P., Conrozier T., Vignon E., Herbage D.,
 RA Mallein-Gerin F.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF247705; AAG22598.1; -; mRNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0005515; F:protein binding; IEA.
 DR GO; GO:0007155; P:cell adhesion; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 DR ProDom; PD000007; C1g_helix; 1.
 KW Collagen.
 FT NON TER 187 187
 FT NON TER 187 187
 SQ SEQUENCE 187 AA; 17467 MW; EBA186A7A69E6D94 CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 187;
 Best Local Similarity 72.7%; Pred. No. 14;
 Matches 8; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

OY 1 GEGKAGSPGL 11
 DB 132 GEGKAGCAPGV 142
 RESULT 67
 Q95J95 CANFA PRELIMINARY; PRT; 194 AA.
 ID Q95J95
 AC Q95J95
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Adiponectin (Fragment).
 GN Name=APM1;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Adipose;
 RA Kabir M., Aanthanarayan S., Ionut V., Kim S.P., Van Citters G.W.,
 RA Dea M.K., Bergman R.N.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF417206; AAL09702.1; -; mRNA.
 DR HSP; Q60994; 1C28.
 DR SMR; Q95J95; 85-194.
 DR Ensembl; ENSCAFG00000013694; Canis familiaris.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR001073; C1g.
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF00386; C1g; 1.
 DR Pfam; PF01391; Collagen; 1.
 DR PRINTS; PR00007; COMPLEMENTC1Q.
 DR ProDom; PD000007; C1g_helix; 1.
 DR SMART; SM00110; C1Q; 1.
 KW Collagen.
 FT NON TER 194 194
 FT NON TER 194 194
 SQ SEQUENCE 194 AA; 20890 MW; 3AA3D947D187AF9A CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 194;
 Best Local Similarity 66.7%; Pred. No. 14;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 OY 1 GEGKAGSPGL 12
 DB 40 GEGKEGDPGLV 51
 RESULT 68
 Q5JPA6 HUMAN PRELIMINARY; PRT; 198 AA.
 ID Q5JPA6
 AC Q5JPA6
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
 DE Hypothetical protein DKFZp434L081.
 GN Name=DKFZp434L081;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
 OC Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC TISSUE=Testis;
 RG The German cDNA Consortium;
 RA Pouetka A., Albert R., Moosmayer P., Schupp I., Wellenreuther R.,

RA Mewes H.W., Weil B., Amid C., Oeanger A., Fobo G., Han M., Wiemann S.;
 RL Submitted (JAN-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL834148; CAI46208.1; -, mRNA.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 DR ProDom; PD000007; Clg_helix; 3.
 KW Collagen; Hypothetical protein.
 SQ SEQUENCE 198 AA; 18217 MW; C98869C0101B7DCE CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 198;
 Best Local Similarity 80.0%; Pred. No. 15;
 Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;
 Qy 1 GEKGAGSPG 10
 Db 170 GEKGPGGAPG 179
 RESULT 69
 Q428Q0 ALOLA
 ID Q428Q0 ALOLA PRELIMINARY; PRT; 229 AA.
 AC Q428Q0;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Adiponectin (Fragment).
 OS Alopec lagopus (Arctic fox).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Alopec.
 OX NCBI_TaxID=9610;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Nieminen P., Hassinen M., Vornanen M., Mustonen A.-M.;
 RT "High homology between the Acrp30 cDNA sequences of wild canids and the domestic dog."
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY965245; AAX73247.1; -, mRNA.
 DR SMR; Q428Q0; 100-229.
 DR InterPro; IPR001073; Clq.
 DR InterPro; IPR008161; Clg_helix.
 DR Pfam; PF00386; Clq; 1.
 DR PRINTS; PR00007; COMPLENTC1Q.
 DR ProDom; PD000007; Clg_helix; 2.
 DR SMART; SM00110; Clq; 1.
 DR PROSITE; PS50871; Clq; 1.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 24588 MW; 2EAF7801C39BF57C CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 229;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPG 10
 Db 170 GEKGPGGAPG 179
 RESULT 70
 Q428Q1 NYCPR
 ID Q428Q1 NYCPR PRELIMINARY; PRT; 229 AA.
 AC Q428Q1;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Adiponectin (Fragment).
 OS Alopec lagopus (Arctic fox).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Alopec.
 OX NCBI_TaxID=9610;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Nieminen P., Hassinen M., Vornanen M., Mustonen A.-M.;
 RT "High homology between the Acrp30 cDNA sequences of wild canids and the domestic dog."
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY965245; AAX73247.1; -, mRNA.
 DR SMR; Q428Q0; 100-229.
 DR InterPro; IPR001073; Clq.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF00386; Clq; 1.
 DR PRINTS; PR00007; Collagen; 1.
 DR ProDom; PD000007; Clg_helix; 2.
 DR SMART; SM00110; Clq; 1.
 DR PROSITE; PS50871; Clq; 1.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 24588 MW; 2EAF7801C39BF57C CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 229;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPG 10
 Db 170 GEKGPGGAPG 179
 RESULT 70
 Q428Q1 NYCPR
 ID Q428Q1 NYCPR PRELIMINARY; PRT; 229 AA.
 AC Q428Q1;
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
 DE Adiponectin (Fragment).
 OS Alopec lagopus (Arctic fox).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Alopec.
 OX NCBI_TaxID=9610;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Nieminen P., Hassinen M., Vornanen M., Mustonen A.-M.;
 RT "High homology between the Acrp30 cDNA sequences of wild canids and the domestic dog."
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY965245; AAX73247.1; -, mRNA.
 DR SMR; Q428Q0; 100-229.
 DR InterPro; IPR001073; Clq.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF00386; Clq; 1.
 DR PRINTS; PR00007; Collagen; 1.
 DR ProDom; PD000007; Clg_helix; 2.
 DR SMART; SM00110; Clq; 1.
 DR PROSITE; PS50871; Clq; 1.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 24588 MW; 2EAF7801C39BF57C CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 229;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPG 10
 Db 170 GEKGPGGAPG 179

OS Nyctereutes procyonoides (Raccoon dog) (Canis procyonoides).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Laurasiatheria; Carnivora; Fissipedia; Canidae;
 OC Nyctereutes.
 OX NCBI_TaxID=34880;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Nieminen P., Hassinen M., Vornanen M., Mustonen A.-M.;
 RT "High homology between the Acrp30 cDNA sequences of wild canids and the domestic dog."
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY965244; AAX73246.1; -, mRNA.
 DR SMR; Q428Q1; 100-229.
 DR InterPro; IPR001073; Clq.
 DR InterPro; IPR008161; Clg_helix.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF00386; Clq; 1.
 DR PRINTS; PR00007; COMPLENTC1Q.
 DR ProDom; PD000007; Clg_helix; 2.
 DR SMART; SM00110; Clq; 1.
 DR PROSITE; PS50871; Clq; 1.
 KW Collagen.
 FT NON_TER 1
 FT NON_TER 229
 SQ SEQUENCE 229 AA; 24588 MW; 2EAF7801C39BF57C CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 229;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPG 10
 Db 55 GEKGPGGAPG 66
 RESULT 71
 Q90612 CHICK
 ID Q90612 CHICK PRELIMINARY; PRT; 310 AA.
 AC Q90612;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein COL3A1 (Fragment).
 GN Name=COL3A1;
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RA Nah H.-D., Niu Z., Adams S.L.;
 RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen."
 RL J. Biol. Chem. 269:16443-16448(1994).
 DR EMBL; U07974; AAA83409.1; -, mRNA.
 DR PUR; I50696; I50696.
 DR Ensembl; ENSGALG0000002552; Gallus gallus.
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0006817; P:phosphate transport; IEA.
 DR InterPro; IPR008160; Collagen.
 DR Pfam; PF01391; Collagen; 4.
 KW Collagen; Hypothetical protein.
 FT NON_TER 310
 SQ SEQUENCE 310 AA; 27601 MW; 5C60B4360832814C CRC64;
 Query Match 74.2%; Score 46; DB 2; Length 310;
 Best Local Similarity 80.0%; Pred. No. 24;
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GEKGAGSPG 10

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CZECH II;
RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC061507; AAH61507.1; -; mRNA.
DR MGI; MGI:188462; Col7al.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR Pfam; PF01391; Collagen; 8.
DR ProDom; PD000007; Clg_helix; 3.
KW Collagen.
FT NON TER
SQ SEQUENCE 546 AA; 51913 MW; 874A0E0431914C65 CRC64;

Query Match 74.2%; Score 46; DB 2; Length 546;
Best Local Similarity 72.7%; Pred. No. 42;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGL 11
|:|:|:|:|
Db 96 GERGAGNGPL 106

RESULT 75
Q99K97 MOUSE
ID Q99K97 MOUSE PRELIMINARY; PRT; 546 AA.
AC Q99K97;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Col4a6 protein (Fragment).
GN Name=Col4a6;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CZECH II;
RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CZECH II;
RC TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004900; AAH04900.1; -; mRNA.
DR HSSP; P08572; 1L11.
DR MGI; MGI:2152695; Col4a6.
DR GO; GO:0005587; C:collagen type IV; IDA.
DR GO; GO:0005615; C:extracellular space; TAS.
DR GO; GO:0005198; F:extracellular molecule activity; IDA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD000007; Clg_helix; 2.
DR ProDom; PD003923; Procollagn4_C; 2.
DR SMART; SM00111; C4; 2.
KW Collagen.
FT NON TER
SQ SEQUENCE 546 AA; 55102 MW; 56F8CC69374BBCE CRC64;

Query Match 74.2%; Score 46; DB 2; Length 546;
Best Local Similarity 72.7%; Pred. No. 42;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGL 11
|:|:|:|:|
Db 81 GKGAGGSPGL 91

Search completed: March 11, 2006, 12:05:50
Job time : 184.8 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:06:07 ; Search time 39.6 Seconds
(without alignments)

Title: US-10-698-121A-2

Perfect score: 62

Sequence: 1 GEKGAEGSPGLL 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Issued Patents AA:*

```
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
```

2: /cgn2_6/ptodata/1/iaa/6_COMB.pcp:*

3: /cgn2_6/ptodata/1/iaa/H_COMB.per:*

4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*

```
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp:"
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	52	83.9	1603	2	US-09-949-016-6136	Sequence 6136, Appli	
2	52	83.9	1609	2	US-09-949-016-10910	Sequence 10910, A	
3	50	80.6	689	2	US-09-949-016-11276	Sequence 11276, A	
4	49	79.0	549	1	US-08-494-168-8	Sequence 8, Appli	
5	49	79.0	1712	2	US-09-961-403-9	Sequence 9, Appli	
6	48	77.4	623	3	US-09-029-348-3	Sequence 3, Appli	
7	48	77.4	626	2	US-09-029-348-2	Sequence 2, Appli	
8	48	77.4	1057	2	US-08-931-820-4	Sequence 4, Appli	
9	48	77.4	1078	2	US-08-963-823-21	Sequence 21, Appli	
10	48	77.4	1078	2	US-09-500-811-21	Sequence 21, Appli	
11	48	77.4	1078	2	US-09-570-573-21	Sequence 21, Appli	
12	48	77.4	1078	2	US-09-548-608-21	Sequence 21, Appli	
13	47	75.8	231	2	US-09-530-423-2	Sequence 2, Appli	
14	47	75.8	244	1	US-08-463-911-7	Sequence 7, Appli	
15	47	75.8	244	2	US-09-140-804-3	Sequence 3, Appli	
16	47	75.8	244	2	US-09-336-536-20	Sequence 20, Appli	
17	47	75.8	244	2	US-09-530-423-1	Sequence 1, Appli	
18	47	75.8	244	2	US-09-686-838B-3	Sequence 3, Appli	
19	47	75.8	244	2	US-09-911-176B-48	Sequence 48, Appli	
20	47	75.8	244	2	US-09-552-225A-3	Sequence 3, Appli	
21	47	75.8	244	2	US-09-619-740-51	Sequence 51, Appli	
22	47	75.8	244	2	US-09-776-976-6	Sequence 6, Appli	
23	47	75.8	244	2	US-09-909-547-6	Sequence 6, Appli	
24	47	75.8	244	2	US-09-569-852B-6	Sequence 6, Appli	
25	47	75.8	244	2	US-09-552-204A-3	Sequence 3, Appli	
26	47	75.8	244	2	US-10-621-787-3	Sequence 3, Appli	
27	47	75.8	1268	2	US-09-949-016-7487	Sequence 7487, Appli	

101	43	69.4	1078	2	US-09-949-016-11185	Sequence 11185, A	174	41	66.1	271	2	US-09-990-444-357	Sequence 357, App
102	43	69.4	1143	2	US-09-949-016-6137	Sequence 6137, Ap	175	41	66.1	271	2	US-09-997-333-357	Sequence 357, App
103	43	69.4	1694	1	US-08-494-168-2	Sequence 2, Appl	176	41	66.1	271	2	US-09-992-598-357	Sequence 357, App
104	42	67.7	62	1	US-07-609-716-61	Sequence 61, Appl	177	41	66.1	468	2	US-09-252-991A-27684	Sequence 27684, A
105	42	67.7	62	2	US-08-475-411A-61	Sequence 61, Appl	178	41	66.1	486	2	US-09-538-092-1269	Sequence 1269, Ap
106	42	67.7	62	2	US-08-478-029A-61	Sequence 61, Appl	179	41	66.1	486	2	US-09-949-016-6151	Sequence 6151, Ap
107	42	67.7	171	2	US-09-011-735-2	Sequence 2, Appl	180	41	66.1	489	1	US-08-794-795-7	Sequence 7, Appl
108	42	67.7	171	2	US-09-029-156-2	Sequence 2, Appl	181	41	66.1	489	1	US-09-249-200-7	Sequence 7, Appl
109	42	67.7	186	2	US-09-366-009-6	Sequence 6, Appl	182	41	66.1	510	2	US-09-949-016-10733	Sequence 10733, A
110	42	67.7	186	2	US-08-809-156B-6	Sequence 6, Appl	183	41	66.1	518	1	US-08-392-367B-2	Sequence 2, Appl
111	42	67.7	186	2	US-09-775-964-6	Sequence 6, Appl	184	41	66.1	518	2	US-08-893-467A-2	Sequence 2, Appl
112	42	67.7	219	2	US-10-153-469A-44	Sequence 44, Appl	185	41	66.1	519	2	US-09-453-702B-265	Sequence 265, App
113	42	67.7	219	2	US-10-153-469A-46	Sequence 46, Appl	186	41	66.1	519	2	US-10-114-170-265	Sequence 265, App
114	42	67.7	219	2	US-10-104-889-44	Sequence 44, Appl	187	41	66.1	1027	2	US-10-360-101-221	Sequence 221, App
115	42	67.7	219	2	US-10-104-889-46	Sequence 46, Appl	188	41	66.1	1218	2	US-09-949-016-7065	Sequence 7065, Ap
116	42	67.7	238	2	US-09-252-991A-16703	Sequence 16703, A	189	40	64.5	36	2	US-09-050-861B-14	Sequence 14, Appl
117	42	67.7	310	2	US-09-219-849-47	Sequence 47, Appl	190	40	64.5	36	2	US-09-135-238B-14	Sequence 14, Appl
118	42	67.7	351	2	US-09-011-735-1	Sequence 1, Appl	191	40	64.5	36	2	US-09-651-150B-14	Sequence 14, Appl
119	42	67.7	351	2	US-09-029-156-1	Sequence 1, Appl	192	40	64.5	163	2	US-10-000-489-58	Sequence 58, Appl
120	42	67.7	464	1	US-08-836-854-19	Sequence 19, Appl	193	40	64.5	177	2	US-10-830-792A-17	Sequence 17, Appl
121	42	67.7	464	2	US-09-366-009-7	Sequence 7, Appl	194	40	64.5	180	2	US-10-830-792A-18	Sequence 18, Appl
122	42	67.7	464	2	US-08-809-156B-7	Sequence 7, Appl	195	40	64.5	186	2	US-10-830-792A-15	Sequence 15, Appl
123	42	67.7	464	2	US-09-775-964-7	Sequence 7, Appl	196	40	64.5	313	2	US-09-949-016-9265	Sequence 9265, Ap
124	42	67.7	489	2	US-09-366-009-8	Sequence 8, Appl	197	40	64.5	313	2	US-10-830-792A-44	Sequence 44, Appl
125	42	67.7	489	2	US-08-809-156B-8	Sequence 8, Appl	198	40	64.5	355	2	US-09-252-991A-27426	Sequence 27426, A
126	42	67.7	489	2	US-09-775-964-8	Sequence 8, Appl	199	40	64.5	599	2	US-09-949-016-8890	Sequence 8890, Ap
127	42	67.7	492	2	US-08-468-996-12	Sequence 12, Appl	200	40	64.5	629	1	US-08-250-740-33	Sequence 33, Appl
128	42	67.7	532	1	US-08-494-168-9	Sequence 9, Appl	201	40	64.5	629	1	US-07-695-472B-2	Sequence 2, Appl
129	42	67.7	546	1	US-08-494-168-10	Sequence 10, Appl	202	40	64.5	629	2	US-09-106-375-2	Sequence 2, Appl
130	42	67.7	684	1	US-09-555-669-12	Sequence 12, Appl	203	40	64.5	638	2	US-10-001-887-108	Sequence 108, App
131	42	67.7	684	2	US-09-073-663-12	Sequence 12, Appl	204	40	64.5	744	2	US-09-949-016-9607	Sequence 9607, Ap
132	42	67.7	833	2	US-09-949-016-9530	Sequence 9530, Ap	205	40	64.5	2149	2	US-09-605-703B-8	Sequence 8, Appl
133	42	67.7	966	2	US-09-949-016-6669	Sequence 6669, Ap	206	39	62.9	21	1	US-07-609-716-50	Sequence 50, Appl
134	42	67.7	1040	2	US-10-153-469A-30	Sequence 30, Appl	207	39	62.9	21	1	US-08-642-255-19	Sequence 19, Appl
135	42	67.7	1040	2	US-10-153-469A-32	Sequence 32, Appl	208	39	62.9	21	2	US-08-475-411A-50	Sequence 50, Appl
136	42	67.7	1040	2	US-10-104-889-30	Sequence 30, Appl	209	39	62.9	21	2	US-09-219-849-1	Sequence 1, Appl
137	42	67.7	1040	2	US-10-104-889-32	Sequence 32, Appl	210	39	62.9	21	2	US-08-478-029A-50	Sequence 50, Appl
138	42	67.7	1366	2	US-09-585-887-10	Sequence 10, Appl	211	39	62.9	62	2	US-08-737-629-10	Sequence 10, Appl
139	42	67.7	1366	2	US-09-289-578-10	Sequence 10, Appl	212	39	62.9	64	2	US-08-737-629-7	Sequence 7, Appl
140	42	67.7	1366	2	US-09-949-016-5882	Sequence 5882, Ap	213	39	62.9	65	2	US-09-513-999C-7428	Sequence 7428, Ap
141	42	67.7	1670	2	US-09-949-016-5883	Sequence 5883, Ap	214	39	62.9	75	2	US-10-830-792A-19	Sequence 19, Appl
142	42	67.7	1690	2	US-09-949-016-5884	Sequence 5884, Ap	215	39	62.9	96	2	US-09-513-999C-4206	Sequence 4206, Ap
143	41	66.1	60	2	US-09-336-536-6	Sequence 6, Appl	216	39	62.9	96	2	US-09-471-276-823	Sequence 823, App
144	41	66.1	60	2	US-09-336-536-13	Sequence 13, Appl	217	39	62.9	198	2	US-09-188-930-138	Sequence 138, App
145	41	66.1	169	2	US-09-706-722A-9	Sequence 9, Appl	218	39	62.9	198	2	US-09-312-283C-138	Sequence 138, App
146	41	66.1	228	2	US-09-336-536-4	Sequence 4, Appl	219	39	62.9	201	2	US-09-902-540-14396	Sequence 14396, A
147	41	66.1	228	2	US-09-336-536-11	Sequence 11, Appl	220	39	62.9	208	2	US-09-800-729-151	Sequence 151, App
148	41	66.1	243	2	US-09-188-930-295	Sequence 295, App	221	39	62.9	245	2	US-09-552-225A-4	Sequence 4, Appl
149	41	66.1	243	2	US-09-140-804-2	Sequence 2, Appl	222	39	62.9	245	2	US-09-552-204A-4	Sequence 4, Appl
150	41	66.1	243	2	US-09-336-536-3	Sequence 3, Appl	223	39	62.9	245	2	US-10-621-787-4	Sequence 4, Appl
151	41	66.1	243	2	US-09-336-536-10	Sequence 10, Appl	224	39	62.9	258	2	US-09-976-594-815	Sequence 815, App
152	41	66.1	243	2	US-09-686-838B-2	Sequence 2, Appl	225	39	62.9	259	2	US-09-991-181-47	Sequence 47, Appl
153	41	66.1	243	2	US-09-312-283C-295	Sequence 295, App	226	39	62.9	259	2	US-09-990-444-47	Sequence 47, Appl
154	41	66.1	243	2	US-09-866-028-42	Sequence 42, Appl	227	39	62.9	259	2	US-09-997-333-47	Sequence 47, Appl
155	41	66.1	243	2	US-09-944-457-42	Sequence 42, Appl	228	39	62.9	259	2	US-09-992-598-47	Sequence 47, Appl
156	41	66.1	243	2	US-09-945-584-42	Sequence 42, Appl	229	39	62.9	260	2	US-09-489-847-198	Sequence 198, App
157	41	66.1	243	2	US-09-944-944-42	Sequence 42, Appl	230	39	62.9	279	2	US-09-010-999-2	Sequence 2, Appl
158	41	66.1	243	2	US-09-945-587-42	Sequence 42, Appl	231	39	62.9	285	2	US-09-312-283C-382	Sequence 382, App
159	41	66.1	247	1	US-08-463-911-2	Sequence 2, Appl	232	39	62.9	285	2	US-09-552-204A-2	Sequence 2, Appl
160	41	66.1	247	2	US-09-140-804-8	Sequence 8, Appl	233	39	62.9	285	2	US-10-621-787-2	Sequence 2, Appl
161	41	66.1	247	2	US-09-118-408-3	Sequence 3, Appl	234	39	62.9	287	2	US-09-489-847-349	Sequence 349, App
162	41	66.1	247	2	US-09-506-855-3	Sequence 3, Appl	235	39	62.9	294	2	US-09-188-930-294	Sequence 294, App
163	41	66.1	247	2	US-09-686-838B-8	Sequence 8, Appl	236	39	62.9	294	2	US-09-312-283C-294	Sequence 294, App
164	41	66.1	247	2	US-09-911-176B-3	Sequence 3, Appl	237	39	62.9	315	2	US-09-602-787A-454	Sequence 454, App
165	41	66.1	247	2	US-09-619-740-3	Sequence 3, Appl	238	39	62.9	375	2	US-09-600-932-29	Sequence 29, Appl
166	41	66.1	247	2	US-09-776-976-2	Sequence 2, Appl	239	39	62.9	430	2	US-09-902-540-13248	Sequence 13248, A
167	41	66.1	247	2	US-09-776-976-4	Sequence 4, Appl	240	39	62.9	446	1	US-08-836-854-15	Sequence 15, Appl
168	41	66.1	247	2	US-09-506-852-3	Sequence 3, Appl	241	39	62.9	484	2	US-09-602-787A-452	Sequence 452, App
169	41	66.1	247	2	US-09-909-547-2	Sequence 2, Appl	242	39	62.9	495	2	US-09-252-991A-31949	Sequence 31949, A
170	41	66.1	247	2	US-09-909-547-4	Sequence 4, Appl	243	39	62.9	571	2	US-10-104-047-3814	Sequence 3814, Ap
171	41	66.1	247	2	US-10-392-706-3	Sequence 3, Appl	244	39	62.9	1024	2	US-08-931-820-2	Sequence 2, Appl
172	41	66.1	251	2	US-09-758-759-194	Sequence 194, App	245	39	62.9	1366	2	US-08-963-825-19	Sequence 19, Appl
173	41	66.1	271	2	US-09-991-181-357	Sequence 357, App	246	39	62.9	1366	2	US-09-500-811-19	Sequence 19, Appl

393	36	58.1	176	2	US-09-252-991A-20852	Sequence 20852, A	466	35	56.5	150	2	US-10-012-896-707	Sequence 707, App
394	36	58.1	182	2	US-09-270-767-42615	Sequence 42615, A	467	35	56.5	159	2	US-09-270-767-32558	Sequence 32558, A
395	36	58.1	212	2	US-09-692-945-2	Sequence 2, Appli	468	35	56.5	159	2	US-09-270-767-47775	Sequence 47775, A
396	36	58.1	242	2	US-08-252-991A-25956	Sequence 25956, A	469	35	56.5	195	2	US-09-252-991A-30084	Sequence 30084, A
397	36	58.1	255	2	US-09-949-016-10619	Sequence 10619, A	470	35	56.5	207	2	US-09-252-991A-26664	Sequence 26664, A
398	36	58.1	273	2	US-08-252-991A-22948	Sequence 22948, A	471	35	56.5	241	2	US-10-076-069-4	Sequence 4, Appli
399	36	58.1	277	2	US-09-600-932-2	Sequence 2, Appli	472	35	56.5	246	2	US-09-552-225A-2	Sequence 2, Appli
400	36	58.1	277	2	US-09-999-833A-97	Sequence 97, Appli	473	35	56.5	246	2	US-10-903-350-14	Sequence 14, Appli
401	36	58.1	277	2	US-10-020-445A-97	Sequence 97, Appli	474	35	56.5	248	2	US-09-198-603C-2	Sequence 2, Appli
402	36	58.1	282	2	US-08-252-991A-19509	Sequence 19509, A	475	35	56.5	248	2	US-09-600-332-27	Sequence 27, Appli
403	36	58.1	317	1	US-08-701-191A-25	Sequence 25, Appli	476	35	56.5	248	2	US-09-949-016-5907	Sequence 5907, Ap
404	36	58.1	317	2	US-09-664-526-25	Sequence 25, Appli	477	35	56.5	248	2	US-09-568-143-1	Sequence 1, Appli
405	36	58.1	376	2	US-09-902-540-12854	Sequence 12854, A	478	35	56.5	251	2	US-09-252-991A-32000	Sequence 32000, A
406	36	58.1	422	1	US-09-252-991A-27513	Sequence 27513, A	479	35	56.5	252	2	US-10-903-350-26	Sequence 26, Appli
407	36	58.1	471	1	US-08-399-889-24	Sequence 24, Appli	480	35	56.5	263	2	US-09-949-016-9370	Sequence 9370, Ap
408	36	58.1	471	2	US-08-167-364-24	Sequence 24, Appli	481	35	56.5	278	2	US-09-252-991A-20931	Sequence 20931, A
409	36	58.1	471	2	US-09-439-897-2	Sequence 2, Appli	482	35	56.5	281	1	US-08-397-633A-75	Sequence 75, Appli
410	36	58.1	482	2	US-09-949-016-9106	Sequence 9106, Ap	483	35	56.5	283	2	US-09-252-991A-21171	Sequence 21171, A
411	36	58.1	484	2	US-09-066-046-8	Sequence 8, Appli	484	35	56.5	288	2	US-08-934-494-4	Sequence 4, Appli
412	36	58.1	501	2	US-08-252-991A-30115	Sequence 30115, A	485	35	56.5	288	2	US-09-143-068-4	Sequence 4, Appli
413	36	58.1	557	2	US-08-979-608A-5	Sequence 5, Appli	486	35	56.5	288	2	US-09-143-707-4	Sequence 4, Appli
414	36	58.1	557	2	US-09-517-849-5	Sequence 5, Appli	487	35	56.5	288	2	US-09-202-089-4	Sequence 4, Appli
415	36	58.1	557	2	US-09-616-289-5	Sequence 5, Appli	488	35	56.5	288	2	US-09-511-133-4	Sequence 4, Appli
416	36	58.1	557	2	US-09-976-740-5	Sequence 5, Appli	489	35	56.5	288	2	US-09-690-169-4	Sequence 4, Appli
417	36	58.1	619	2	US-09-066-046-2	Sequence 2, Appli	490	35	56.5	288	2	US-09-511-631-4	Sequence 4, Appli
418	36	58.1	633	2	US-08-581-001B-12	Sequence 12, Appli	491	35	56.5	288	2	US-09-690-189-4	Sequence 4, Appli
419	36	58.1	647	2	US-08-452-846E-4	Sequence 4, Appli	492	35	56.5	290	1	US-08-440-846-2	Sequence 2, Appli
420	36	58.1	716	2	US-09-252-991A-30683	Sequence 30683, A	493	35	56.5	309	2	US-10-903-350-24	Sequence 24, Appli
421	36	58.1	722	2	US-09-252-991A-24102	Sequence 24102, A	494	35	56.5	315	2	US-09-270-767-44966	Sequence 44966, A
422	36	58.1	734	2	US-08-252-991A-30703	Sequence 30703, A	495	35	56.5	316	2	US-09-902-540-15157	Sequence 15157, A
423	36	58.1	738	2	US-09-252-991A-30721	Sequence 30721, A	496	35	56.5	319	2	US-10-012-605C-2	Sequence 2, Appli
424	36	58.1	825	2	US-08-540-824-26	Sequence 26, Appli	497	35	56.5	319	2	US-09-758-759-89	Sequence 89, Appli
425	36	58.1	875	2	US-08-252-991A-30056	Sequence 30056, A	498	35	56.5	320	2	US-09-949-016-6177	Sequence 6177, Ap
426	36	58.1	913	1	US-08-445-640-4	Sequence 4, Appli	499	35	56.5	320	2	US-09-949-016-10451	Sequence 10451, A
427	36	58.1	913	2	US-08-170-558-4	Sequence 4, Appli	500	35	56.5	336	2	US-09-252-991A-16971	Sequence 16971, A
428	36	58.1	913	2	US-08-447-314-4	Sequence 4, Appli							
429	36	58.1	913	2	US-08-445-461-4	Sequence 4, Appli							
430	36	58.1	913	2	US-09-140-378A-2	Sequence 2, Appli							
431	36	58.1	913	2	US-08-223-490-4	Sequence 4, Appli							
432	36	58.1	919	1	US-08-336-343A-2	Sequence 2, Appli							
433	36	58.1	919	2	US-09-551-188-2	Sequence 2, Appli							
434	36	58.1	937	2	US-09-949-016-8366	Sequence 8366, Ap							
435	36	58.1	937	2	US-09-949-016-8367	Sequence 8367, Ap							
436	36	58.1	937	2	US-09-949-016-8368	Sequence 8368, Ap							
437	36	58.1	970	2	US-08-134-000C-4256	Sequence 4256, Ap							
438	36	58.1	1428	2	US-08-644-827B-7	Sequence 7, Appli							
439	36	58.1	2104	1	US-08-808-793-4	Sequence 4, Appli							
440	36	58.1	2104	2	US-08-772-512A-4	Sequence 4, Appli							
441	36	58.1	2104	2	US-08-428-371-4	Sequence 4, Appli							
442	36	58.1	2105	1	US-08-808-793-3	Sequence 3, Appli							
443	36	58.1	2105	2	US-08-772-512A-3	Sequence 3, Appli							
444	36	58.1	2105	2	US-08-428-371-3	Sequence 3, Appli							
445	36	58.1	2647	1	US-08-583-562B-8	Sequence 8, Appli							
446	36	58.1	2647	1	US-08-779-113-8	Sequence 8, Appli							
447	36	58.1	2647	2	US-09-949-016-6082	Sequence 6082, Ap							
448	36	58.1	2666	2	US-08-949-016-10857	Sequence 10857, A							
449	35	56.5	9	2	US-08-159-335A-821	Sequence 821, App							
450	35	56.5	47	2	US-08-851-873-45	Sequence 45, Appli							
451	35	56.5	53	1	US-08-316-650-10	Sequence 10, Appli							
452	35	56.5	53	2	US-08-479-722B-10	Sequence 10, Appli							
453	35	56.5	53	2	US-09-592-685-10	Sequence 10, Appli							
454	35	56.5	53	4	PCT-US95-02251-10	Sequence 10, Appli							
455	35	56.5	90	2	US-10-830-792A-22	Sequence 22, Appli							
456	35	56.5	101	2	US-09-621-976-4945	Sequence 4945, Ap							
457	35	56.5	102	2	US-09-471-276-1560	Sequence 1560, Ap							
458	35	56.5	117	2	US-10-076-069-7	Sequence 7, Appli							
459	35	56.5	149	2	US-09-252-991A-20244	Sequence 20244, A							
460	35	56.5	150	2	US-08-636-215-707	Sequence 707, App							
461	35	56.5	150	2	US-08-685-166A-707	Sequence 707, App							
462	35	56.5	150	2	US-09-679-426-707	Sequence 707, App							
463	35	56.5	150	2	US-09-759-143-707	Sequence 707, App							
464	35	56.5	150	2	US-09-651-236-707	Sequence 707, App							
465	35	56.5	150	2	US-09-657-279-707	Sequence 707, App							

ALIGNMENTS

RESULT 1

US-09-949-016-6136
; Sequence 6136, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6136
; LENGTH: 1603
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-6136

Query Match 83.9%; Score 52; DB 2; Length 1603;
Best Local Similarity 81.8%; Pred. No. 7.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11

Db 817 GKGAGSGPGL 827

Query Match 79.0%; Score 49; DB 2; Length 1712;

```
Best Local Similarity 72.7%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGL 11
Db 276 GKGSGSGPGI 286

RESULT 6
US-09-029-348-3
; Sequence 3, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-3

Query Match 77.4%; Score 48; DB 2; Length 623;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
Db 330 GERGSGSGSPG 339

RESULT 7
US-09-029-348-2
; Sequence 2, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: G087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM CDNA OF PROCOLLAGENS
US-09-029-348-2

Query Match 77.4%; Score 48; DB 2; Length 626;
Best Local Similarity 80.0%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
Db 330 GERGSGSGSPG 339

RESULT 8
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
```

```
APPLICANT:
TITLE OF INVENTION: Assay for collagen degradation
NUMBER OF SEQUENCES: 4
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/931,820
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 96202596.1
FILING DATE:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Collagen type III
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1055
OTHER INFORMATION: /label= Modified
OTHER INFORMATION: /note= "Ala may be Pro"
US-08-931-820-4

Query Match 77.4%; Score 48; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
Db 1022 GERGSGSGPG 1031

RESULT 9
US-08-963-825-21
; Sequence 21, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
```

ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1078 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (III)
 US-08-963-825-21

Query Match 77.4%; Score 48; DB 2; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
 |||:|||||
 Db 1023 GERSESGSPG 1032

RESULT 10
 US-09-500-811-21
 ; Sequence 21, Application US/09500811
 ; Patent No. 6323314
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/500,811
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 21:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1078 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear

MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN ALPHA 1 (III)
 US-09-500-811-21

Query Match 77.4%; Score 48; DB 2; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
 |||:|||||
 Db 1023 GERSESGSPG 1032

RESULT 11
 US-09-570-573-21
 ; Sequence 21, Application US/09570573
 ; Patent No. 6342361
 ; GENERAL INFORMATION:
 ; APPLICANT: Qvist, Per
 ; APPLICANT: Bonde, Martin
 ; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
 ; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
 ; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
 ; TITLE OF INVENTION: Disorders Associated with the Metabolism of
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby & Darby PC
 ; STREET: 805 Third Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/570,573
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/187,319
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 21:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1078 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (III)
 ; US-09-570-573-21

Query Match 77.4%; Score 48; DB 2; Length 1078;
 Best Local Similarity 80.0%; Pred. No. 22;
 Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
 |||:|||||

Db 1023 GERGEGSPG 1032

RESULT 12

US-09-548-608-21

; Sequence 21, Application US/09548608

; Patent No. 6355442

; GENERAL INFORMATION:

; APPLICANT: Qvist, Per

; APPLICANT: Bonde, Martin

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments

; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the

; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

; TITLE OF INVENTION: Disorders Associated with the Metabolism of

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Darby & Darby PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/548,608

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/187,319

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Gogoris, Adga C

; REGISTRATION NUMBER: 29,714

; REFERENCE/DOCKET NUMBER: 4305/08701

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 21:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1078 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; ORIGINAL SOURCE:

; ORGANISM: Homo sapiens

; IMMEDIATE SOURCE:

; CLONE: COLLAGEN ALPHA 1 (III)

US-09-548-608-21

Query Match 77.4%; Score 48; DB 2; Length 1078;

Best Local Similarity 80.0%; Pred. No. 22;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GERGEGSPG 10

||:|||||

Db 1023 GERGEGSPG 1032

RESULT 13

US-09-530-423-2

; Sequence 2, Application US/09530423

; Patent No. 6461821

; GENERAL INFORMATION:

; APPLICANT: Otsuka Pharmaceutical Co., Ltd.

; TITLE OF INVENTION: Smooth muscle growth inhibitory composition, a

; TITLE OF INVENTION: diagnostic method for arteriosclerosis and a kit

; TITLE OF INVENTION: therefor

; FILE REFERENCE: P98-51

; Sequence 3, Application US/09140804
; Patent No. 6197930
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOGS
; FILE REFERENCE: 97-49
; CURRENT APPLICATION NUMBER: US/09/140,804
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: 60/056,983
; EARLIER FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-140-804-3

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
| | | | : | | | |
Db 63 GEGKAGSPGLI 74

RESULT 16
US-09-336-536-20
; Sequence 20, Application US/09336536
; Patent No. 6406884
; GENERAL INFORMATION:
; APPLICANT: Leiby, K.
; APPLICANT: McKay, C.
; APPLICANT: Bossone, S.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-144
; CURRENT APPLICATION NUMBER: US/09/336,536
; CURRENT FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-336-536-20

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
| | | | : | | | |
Db 63 GEGKAGSPGLI 74

RESULT 17
US-09-530-423-1
; Sequence 1, Application US/09530423
; Patent No. 6461821
; GENERAL INFORMATION:
; APPLICANT: Otsuka Pharmaceutical Co., Ltd.
; TITLE OF INVENTION: Smooth muscle growth inhibitory composition, a
; TITLE OF INVENTION: diagnostic method for arteriosclerosis and a kit
; TITLE OF INVENTION: therefor
; FILE REFERENCE: P98-51
; CURRENT APPLICATION NUMBER: US/09/530,423
; CURRENT FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: JP H9-297569
; PRIOR FILING DATE: 1997-10-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Abdominal fat tissue from myoma uteri
US-09-530-423-1

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
| | | | : | | | |
Db 63 GEGKAGSPGLI 74

RESULT 18
US-09-686-838B-3
; Sequence 3, Application US/09686838B
; Patent No. 6482612
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Humes, Jacqueline M.
; TITLE OF INVENTION: Adipocyte-Specific Protein Homologs
; FILE REFERENCE: 97-49D1
; CURRENT APPLICATION NUMBER: US/09/686,838B
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/140,804
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: US 60/056,983
; PRIOR FILING DATE: 1997-08-26
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-686-838B-3

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
| | | | : | | | |
Db 63 GEGKAGSPGLI 74

RESULT 19
US-09-911-176B-48
; Sequence 48, Application US/09911176B
; Patent No. 6518403
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ANTIBODIES THAT BIND AN
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG
; FILE REFERENCE: 97-30D1
; CURRENT APPLICATION NUMBER: US/09/911,176B
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/118,408
; PRIOR FILING DATE: 1998-07-17
; PRIOR APPLICATION NUMBER: 60/053,154
; PRIOR FILING DATE: 1997-07-18
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 48
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-911-176B-48

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
||||:||||:
Db 63 GEGKEGDPGLI 74

RESULT 20

US-09-552-225A-3
; Sequence 3, Application US/09552225A
; Patent No. 6521233
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Bishop, Paul
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG ZACRP3
; FILE REFERENCE: 99-09
; CURRENT APPLICATION NUMBER: US/09/552.225A
; CURRENT FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/130,199
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-552-225A-3

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
||||:||||:
Db 63 GEGKEGDPGLI 74

RESULT 21

US-09-619-740-51
; Sequence 51, Application US/09619740
; Patent No. 6544946
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Lasser, Gerald W.
; APPLICANT: Bishop, Paul D.
; TITLE OF INVENTION: INHIBITORS FOR USE IN HEMOSTASIS AND IMMUNE FUNCTION
; FILE REFERENCE: 99-12C3
; CURRENT APPLICATION NUMBER: US/09/619,740
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/253,604
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: 09/444,794
; PRIOR FILING DATE: 1999-11-22
; PRIOR APPLICATION NUMBER: 09/506,855
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 51
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-619-740-51

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
||||:||||:
Db 63 GEGKEGDPGLI 74

RESULT 22

US-09-776-976-6

; Sequence 6, Application US/09776976
; Patent No. 6566332
; GENERAL INFORMATION:
; APPLICANT: Fruebis, Joachim
; APPLICANT: Erickson, Mary Ruth
; APPLICANT: Yen, Frances
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
; FILE REFERENCE: 76.US4.REG
; CURRENT APPLICATION NUMBER: US/09/776,976
; CURRENT FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/758,055
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 60/176,228
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/198,087
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 60/299,881
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 6
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-776-976-6

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
||||:||||:
Db 63 GEGKEGDPGLI 74

RESULT 23

US-09-909-547-6
; Sequence 6, Application US/09909547
; Patent No. 6579852
; GENERAL INFORMATION:
; APPLICANT: Fruebis, Joachim
; APPLICANT: Erickson, Mary Ruth
; APPLICANT: Yen, Frances
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: OBG3 Globular Head and Uses Thereof for Decreasing Body Mass
; FILE REFERENCE: 76.US6.CIP
; CURRENT APPLICATION NUMBER: US/09/909,547
; CURRENT FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: US 09/776,976
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/758,055
; PRIOR FILING DATE: 2001-01-10
; PRIOR APPLICATION NUMBER: US 60/299,881
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 60/198,087
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: US 60/176,228
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 6
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-909-547-6

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSPGLL 12
||||:||||:
Db 63 GEGKEGDPGLI 74

Db 63 GEKGEKDPGLI 74

US-09-569-852B-6

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 24

US-09-569-852B-6

Sequence 6, Application US/09569852B

Patent No. 6582909

GENERAL INFORMATION:

APPLICANT: Bouqueleret, Lydie

APPLICANT: Bihain, Bernard

APPLICANT: Denison, Blake

APPLICANT: Yen-Potin, Frances

TITLE OF INVENTION: APM1 Biallelic Markers and Uses Thereof

FILE REFERENCE: GEN-T113XC2

CURRENT APPLICATION NUMBER: US/09/569,852B

CURRENT FILING DATE: 2002-03-12

PRIOR FILING DATE: 1999-11-04

PRIOR APPLICATION NUMBER: US 09/434,848

PRIOR FILING DATE: 1999-11-04

PRIOR APPLICATION NUMBER: US 60/119,593

PRIOR FILING DATE: 1999-02-10

PRIOR APPLICATION NUMBER: US 60/107,113

PRIOR FILING DATE: 1998-11-04

NUMBER OF SEQ ID NOS: 8

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 244

TYPE: PRT

ORGANISM: Homo sapiens

NAME/KEY: misc.feature

LOCATION: (15)..(15)

OTHER INFORMATION: The 'Xaa' at location 15 stands for Gly.

NAME/KEY: misc.feature

LOCATION: (1)..(367)

OTHER INFORMATION: homology with 5' EST A254990 in private bank : GENSET

NAME/KEY: misc.feature

LOCATION: (91)..(93)

OTHER INFORMATION: Amino acid at position 15 (Xaa) means Gly

NAME/KEY: misc.feature

LOCATION: (15)..(15)

OTHER INFORMATION: The 'Xaa' at location 15 stands for Gly.

US-09-569-852B-6

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGEKSPGLL 12

Db 63 GEKGEKDPGLI 74

US-09-552-204A-3

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 25

US-09-552-204A-3

Sequence 3, Application US/09552204A

Patent No. 6620909

GENERAL INFORMATION:

APPLICANT: Piddington, Christopher S.

APPLICANT: Bishop, Paul D.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG ZACRP2

FILE REFERENCE: 99-08

CURRENT APPLICATION NUMBER: US/09/552,204A

CURRENT FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: 60/130,207

PRIOR FILING DATE: 1999-04-20

NUMBER OF SEQ ID NOS: 20

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 244

TYPE: PRT

ORGANISM: Homo sapiens

US-09-949-016-7487

Query Match 75.8%; Score 47; DB 2; Length 1268;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGEKSPGLL 12

Db 63 GEKGEKDPGLI 74

US-09-949-016-7487

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 26

US-10-621-787-3

Sequence 3, Application US/10621787

Patent No. 6921649

GENERAL INFORMATION:

APPLICANT: Piddington, Christopher S.

APPLICANT: Bishop, Paul D.

TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG

FILE REFERENCE: 99-08D1

CURRENT APPLICATION NUMBER: US/10/621,787

CURRENT FILING DATE: 2003-07-17

PRIOR APPLICATION NUMBER: US 09/552,204

PRIOR FILING DATE: 2000-04-19

PRIOR APPLICATION NUMBER: US 60/130,207

PRIOR FILING DATE: 1999-04-20

NUMBER OF SEQ ID NOS: 23

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 244

TYPE: PRT

ORGANISM: Homo sapiens

US-10-621-787-3

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGEKSPGLL 12

Db 63 GEKGEKDPGLI 74

US-09-949-016-7487

Query Match 75.8%; Score 47; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 7.2;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

RESULT 27

US-09-949-016-7487

Sequence 7487, Application US/09949016

Patent No. 6812339

GENERAL INFORMATION:

APPLICANT: VENTER, J. Craig et al.

TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

FILE REFERENCE: CL001307

CURRENT APPLICATION NUMBER: US/09/949,016

CURRENT FILING DATE: 2000-04-14

PRIOR APPLICATION NUMBER: 60/241,755

PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/237,768

PRIOR FILING DATE: 2000-10-03

PRIOR APPLICATION NUMBER: 60/231,498

PRIOR FILING DATE: 2000-09-08

NUMBER OF SEQ ID NOS: 207012

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7487

LENGTH: 1268

TYPE: PRT

ORGANISM: Human

US-09-949-016-7487

Query Match 75.8%; Score 47; DB 2; Length 1268;
Best Local Similarity 72.7%; Pred. No. 36;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPGL 11
||||:||||
Db 938 GEGSKGEPGL 948

RESULT 28

US-09-640-211A-863
; Sequence 863, Application US/09640211A
; Patent No. 6833446

; GENERAL INFORMATION:

; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021CIU
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 863
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Eucalyptus grandis
US-09-640-211A-863

Query Match 74.2%; Score 46; DB 2; Length 182;
Best Local Similarity 66.7%; Pred. No. 7.7;
Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPGL 12
||||:||||
Db 25 GEGGADGVPGL 36

RESULT 29

US-09-623-497-1
; Sequence 1, Application US/09623497
; Patent No. 6706490

; GENERAL INFORMATION:

; APPLICANT: COOK, ANDREW
; APPLICANT: ROWLEY, MERRILL
; APPLICANT: MACKAY, IAN
; TITLE OF INVENTION: METHOD FOR THE DIAGNOSIS OF RHEUMATOID ARTHRITIS
; FILE REFERENCE: 017227/0167
; CURRENT APPLICATION NUMBER: US/09/623,497
; CURRENT FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: PCT/AU98/00176
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: AU POS712/97
; PRIOR FILING DATE: 1997-03-19
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 347
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-623-497-1

Query Match 74.2%; Score 46; DB 2; Length 347;
Best Local Similarity 80.0%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||||:||||
Db 21 GEGGPEGAPG 30

RESULT 30

US-08-468-996-11
; Sequence 11, Application US/08468996
; Patent No. 6645504

; GENERAL INFORMATION:

; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION OF
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 492
; TYPE: PRT
; ORGANISM: Bos taurus
US-08-468-996-11

Query Match 74.2%; Score 46; DB 2; Length 492;
Best Local Similarity 80.0%; Pred. No. 20;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||||:||||
Db 451 GEGGPEGAPG 460

RESULT 31

US-09-252-991A-32551
; Sequence 32551, Application US/09252991A
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32551
; LENGTH: 926
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32551

Query Match 74.2%; Score 46; DB 2; Length 926;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GKGAGSPG 10
: : : : :
Db 872 GDSGAGSPG 881

RESULT 32

US-08-468-996-10.
; Sequence 10, Application US/08468996
; Patent No. 6645504
; GENERAL INFORMATION:
; APPLICANT: Weiner, Howard
; APPLICANT: Miller, Ariel
; APPLICANT: Zheng, Zheng
; APPLICANT: Ahmad, Al-Sabbagh
; TITLE OF INVENTION: BYSTANDER SUPPRESSION OF TYPE 1 DIABETES BY ORAL ADMINISTRATION C
; TITLE OF INVENTION: GLUCAGON
; FILE REFERENCE: 1010/16959-US3
; CURRENT APPLICATION NUMBER: US/08/468,996
; CURRENT FILING DATE: 2003-02-07
; PRIOR APPLICATION NUMBER: US 07/843,752
; PRIOR FILING DATE: 1992-02-28
; PRIOR APPLICATION NUMBER: US 07/460,852
; PRIOR FILING DATE: 1990-02-21
; PRIOR APPLICATION NUMBER: US 07/596,936
; PRIOR FILING DATE: 1990-10-15
; PRIOR APPLICATION NUMBER: US 07/065,734
; PRIOR FILING DATE: 1987-06-24
; PRIOR APPLICATION NUMBER: US 07/454,486
; PRIOR FILING DATE: 1989-12-20
; PRIOR APPLICATION NUMBER: US 07/487,732
; PRIOR FILING DATE: 1990-03-02
; PRIOR APPLICATION NUMBER: US 07/551,632
; PRIOR FILING DATE: 1990-07-10
; PRIOR APPLICATION NUMBER: US 07/379,778
; PRIOR FILING DATE: 1989-07-14
; PRIOR APPLICATION NUMBER: US 07/607,826
; PRIOR FILING DATE: 1990-10-31
; PRIOR APPLICATION NUMBER: US 07/595,468
; PRIOR FILING DATE: 1990-10-10
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 10
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-468-996-10

Query Match 74.2%; Score 46; DB 2; Length 1017;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GKGAGSPG 10
: : : : :
Db 571 GKGPEGAPG 580

RESULT 33

US-08-931-820-3
; Sequence 3, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type II
US-08-931-820-3

Query Match 74.2%; Score 46; DB 2; Length 1060;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GKGAGSPG 10
: : : : :
Db 590 GKGPEGAPG 599

RESULT 34

US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689

; GENERAL INFORMATION:
; APPLICANT: Qvist, Per

; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:

; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Cogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:

CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match 74.2%; Score 46; DB 2; Length 1418;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 702 GEGGEGAPG 711

RESULT 35

US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
; GENERAL INFORMATION:
; APPLICANT: Poole, Anthony R.
; APPLICANT: Hollander, Anthony P.
; APPLICANT: Billinghamurst, R. C.
; TITLE OF INVENTION: IMMUNOSAYS FOR THE MEASUREMENT OF
; TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,999
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 4335
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/448,501
; FILING DATE: 17-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,123
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 032931/0212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Human Type II Collagen
US-09-010-999-1

Query Match 74.2%; Score 46; DB 2; Length 1418;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 702 GEGGEGAPG 711

RESULT 36

US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C.
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-500-811-20

Query Match 74.2%; Score 46; DB 2; Length 1418;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 702 GEGGEGAPG 711

RESULT 37

US-09-570-573-20
; Sequence 20, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:

FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 FILING DATE:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Gogoris, Adda C
 REGISTRATION NUMBER: 29,714
 REFERENCE/DOCKET NUMBER: 4305/08701
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-527-7700
 TELEFAX: 212-753-6237
 TELEX: 236687
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1418 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: COLLAGEN -ALPHA 1 (II)
 US-09-548-608-20

Query Match 74.2%; Score 46; DB 2; Length 1418;
 Best Local Similarity 80.0%; Pred.No.58;
 Matches 8; Conservative 1; Mismatches 1; Indels

QY 1 GKGAGSGPG 10
 |||||
 Db 702 GKGPGGAPG 711

RESULT 39
 US-08-316-650-12
 ; Sequence 12, Application US/08316650
 ; Patent No. 5942496
 ; GENETIC INFORMATION:
 ; APPLICANT: Bonadio, Jeffrey
 ; APPLICANT: Roessler, Blake J.
 ; APPLICANT: Goldstein, Steven A.
 ; APPLICANT: Lin, Wushan
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS
 ; TITLE OF INVENTION: FOR STIMULATING BONE CELLS
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/316,650
 ; FILING DATE: 30-SEP-1994
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/199,780
 ; FILING DATE: 30-SEP-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Parker, David L.
 ; REGISTRATION NUMBER: 32,165
 ; REFERENCE/DOCKET NUMBER: UMIC:008
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (512) 418-3000
 ; TELEFAX: (713) 789-2679
 ; TELEX: 79-0924

; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-316-650-12

Query Match 74.2%; Score 46; DB 1; Length 1442;
Best Local Similarity 80.0%; Pred. No. 59;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
Db 726 GEKGPEGAPG 735
|||||:|

RESULT 40
PCT-US95-02251-12
; Sequence 12, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT: METHODS AND COMPOSITIONS FOR STIMULATING BONE
; TITLE OF INVENTION: CELLS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02251
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC009P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US95-02251-12

Query Match 74.2%; Score 46; DB 4; Length 1442;
Best Local Similarity 80.0%; Pred. No. 59;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
Db 726 GEKGPEGAPG 735
|||||:|

Db 726 GEKGPEGAPG 735

RESULT 41
US-09-795-061-2
; Sequence 2, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Inamura, Yasutada
; TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-795-061-2

Query Match 74.2%; Score 46; DB 2; Length 1739;
Best Local Similarity 80.0%; Pred. No. 71;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
Db 691 GEKGAGGPPG 700
|||||:|

RESULT 42
US-09-795-061-4
; Sequence 4, Application US/09795061
; Patent No. 6759528
; GENERAL INFORMATION:
; APPLICANT: Greenspan, Daniel S
; APPLICANT: Inamura, Yasutada
; TITLE OF INVENTION: Pro-Alpha 3 (V) Collagen Genes
; FILE REFERENCE: 960296.96781
; CURRENT APPLICATION NUMBER: US/09/795,061
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-795-061-4

Query Match 74.2%; Score 46; DB 2; Length 1745;
Best Local Similarity 80.0%; Pred. No. 71;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GEKGAGSGPG 10
Db 581 GEKGAGGPPG 590
|||||:|

RESULT 43
US-09-949-002-405
; Sequence 405, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08

```
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-405

Query Match      74.2%; Score 46; DB 2; Length 1745;
Best Local Similarity 80.0%; Pred. No. 71;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GKGAGGSPG 10
Db      581 GKGAGGPPG 590

RESULT 44
US-09-949-002-492
; Sequence 492, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 492
; LENGTH: 1771
; TYPE: PRT
; ORGANISM: Human
US-09-949-002-492

Query Match      74.2%; Score 46; DB 2; Length 1771;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GKGAGGSPG 10
Db      607 GKGAGGPPG 616

RESULT 45
US-09-919-497-56
; Sequence 56, Application US/09919497
; Patent No. 6773883
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/7225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: xaa = any amino acid
```

```
US-09-919-497-56

Query Match      74.2%; Score 46; DB 2; Length 1806;
Best Local Similarity 72.7%; Pred. No. 73;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GKGAGGSPGL 11
Db      712 GKGPGQKPGGL 722

RESULT 46
US-08-642-255-31
; Sequence 31, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; PROTEIN POLYMERS
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, ROHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 PHT UR
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 62 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-642-255-31

Query Match      72.6%; Score 45; DB 1; Length 62;
Best Local Similarity 80.0%; Pred. No. 3.8;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GKGAGGSPG 10
Db      43 GPKGADGSPG 52

RESULT 47
US-07-609-716-111
; Sequence 111, Application US/07609716
; Patent No. 5514581
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; SYNTHETIC PROTEIN POLYMER
; NUMBER OF SEQUENCES: 118
```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;; STREET: Four Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: CA
;; COUNTRY: US
;; ZIP: 94111
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/609,716
;; FILING DATE: 06-NOV-1990
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rowland, Bertram I
;; REGISTRATION NUMBER: 20015
;; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
;; TELEPHONE: 415-781-1989
;; TELEFAX: 415-398-3249
;; INFORMATION FOR SEQ ID NO: 111:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 69 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-07-609-716-111

Query Match 72.6%; Score 45; DB 1; Length 69;
Best Local Similarity 80.0%; Pred. No. 4.3;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

Qy 1 GEKGAGSGPG 10
| | | | |
Db 40 GPKGADGSPG 49

RESULT 48
US-08-475-411A-111
; Sequence 111, Application US/08475411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/114,618
;; FILING DATE: 29-OCT-1987
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 06/927,258
;; FILING DATE: 04-NOV-1986
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Trecartin, Richard F.
;; REGISTRATION NUMBER: 31,801
;; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-781-1989
;; TELEFAX: 415-398-3249
;; INFORMATION FOR SEQ ID NO: 111:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 69 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-475-411A-111

Query Match 72.6%; Score 45; DB 2; Length 69;
Best Local Similarity 80.0%; Pred. No. 4.3;
Matches 8; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

Qy 1 GEKGAGSGPG 10
| | | | |
Db 40 GPKGADGSPG 49

RESULT 49
US-08-478-029A-111
; Sequence 111, Application US/08478029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK

TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-478-029A-111

Query Match 72.6%; Score 45; DB 2; Length 69;
Best Local Similarity 80.0%; Pred. No. 4.3;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
DB 40 GPKGADGSGPG 49

RESULT 50
US-07-609-716-113
Sequence 113, Application US/07609716
Patent No. 5514581
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 72 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-07-609-716-113

Query Match 72.6%; Score 45; DB 1; Length 72;
Best Local Similarity 80.0%; Pred. No. 4.4;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
DB 40 GPKGADGSGPG 49

RESULT 51

US-08-475-411A-113
Sequence 113, Application US/08475411A
Patent No. 6140072
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,411A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 72 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-411A-113

Query Match 72.6%; Score 45; DB 2; Length 72;
Best Local Similarity 80.0%; Pred. No. 4.4;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
DB 40 GPKGADGSGPG 49

RESULT 52
US-08-478-029A-113
Sequence 113, Application US/08478029A
Patent No. 6184348
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;; STREET: Four Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: CA
;; COUNTRY: US
;; ZIP: 94111
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/478,029A
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/609,716
;; FILING DATE: 06-NOV-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/269,429
;; FILING DATE: 09-NOV-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 06/927,258
;; FILING DATE: 04-NOV-1986
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Trecartin, Richard F.
;; REGISTRATION NUMBER: 31,801
;; REFERENCE/DOCKET NUMBER: A-55186-8/RPT/MTK
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-781-1989
;; TELEFAX: 415-398-3249
;; INFORMATION FOR SEQ ID NO: 113:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 72 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-08-478-029A-113

Query Match 72.6%; Score 45; DB 2; Length 72;
Best Local Similarity 80.0%; Pred. No. 4.4;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPG 10
Db 40 GPRGADGSPG 49

RESULT 53
US-07-609-716-114
;; Sequence 114, Application US/07609716
;; Patent No. 5514581
;; GENERAL INFORMATION:
;; APPLICANT: Ferrari, Franco A.
;; APPLICANT: Cappello, Joseph
;; TITLE OF INVENTION: Functional Recombinantly Prepared
;; TITLE OF INVENTION: Synthetic Protein Polymer
;; NUMBER OF SEQUENCES: 118
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;; STREET: Four Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: CA
;; COUNTRY: US
;; ZIP: 94111
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/07/609,716
;; FILING DATE: 06-NOV-1990
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rowland, Bertram I
;; REGISTRATION NUMBER: 20015
;; REFERENCE/DOCKET NUMBER: A-55186-3/BIR
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-781-1989
;; TELEFAX: 415-398-3249
;; INFORMATION FOR SEQ ID NO: 114:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 82 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; US-07-609-716-114

Query Match 72.6%; Score 45; DB 1; Length 82;
Best Local Similarity 80.0%; Pred. No. 5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPG 10
Db 40 GPRGADGSPG 49

RESULT 54
US-08-475-411A-114
;; Sequence 114, Application US/08475411A
;; Patent No. 6140072
;; GENERAL INFORMATION:
;; APPLICANT: Ferrari, Franco A.
;; APPLICANT: Cappello, Joseph
;; TITLE OF INVENTION: Functional Recombinantly Prepared
;; TITLE OF INVENTION: Synthetic Protein Polymer
;; NUMBER OF SEQUENCES: 119
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;; STREET: Four Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: CA
;; COUNTRY: US
;; ZIP: 94111
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/475,411A
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/609,716
;; FILING DATE: 06-NOV-1990
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/269,429
;; FILING DATE: 09-NOV-1988
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/114,618
;; FILING DATE: 29-OCT-1987
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 06/927,258
;; FILING DATE: 04-NOV-1986
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Trecartin, Richard F.
;; REGISTRATION NUMBER: 31,801
;; REFERENCE/DOCKET NUMBER: A-55186-9/RPT/MTK
;; TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-475-411A-114

Query Match 72.6%; Score 45; DB 2; Length 85;
Best Local Similarity 80.0%; Pred. No. 5.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 43 GPKGADGSPG 52

RESULT 55

US-08-478-029A-114
Sequence 114, Application US/08478029A
Patent No. 6184348
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,029A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/609,716
FILING DATE: 06-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/269,429
FILING DATE: 09-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/114,618
FILING DATE: 29-OCT-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 06/927,258
FILING DATE: 04-NOV-1986
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 114:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-478-029A-114

Query Match 72.6%; Score 45; DB 2; Length 85;
Best Local Similarity 80.0%; Pred. No. 5.2;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 43 GPKGADGSPG 52

RESULT 56

US-07-609-716-66
Sequence 66, Application US/07609716
Patent No. 5514581
GENERAL INFORMATION:
APPLICANT: Ferrari, Franco A.
APPLICANT: Cappello, Joseph
TITLE OF INVENTION: Functional Recombinantly Prepared
TITLE OF INVENTION: Synthetic Protein Polymer
NUMBER OF SEQUENCES: 118
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/609,716
FILING DATE: 06-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A-55186-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 357 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-609-716-66

Query Match 72.6%; Score 45; DB 1; Length 357;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 76 GPKGADGSPG 85

RESULT 57

US-08-642-255-33
Sequence 33, Application US/08642255
Patent No. 5773249
GENERAL INFORMATION:
APPLICANT: Cappello, Joseph
APPLICANT: Ferrari, Franco A.
TITLE OF INVENTION: High Molecular Weight Collagen-Like
TITLE OF INVENTION: Protein Polymers
NUMBER OF SEQUENCES: 135
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

```
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-33

Query Match 72.6%; Score 45; DB 1; Length 357;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
Db 76 GPKGADGSGPG 85

RESULT 58
US-08-475-411A-66
; Sequence 66, Application US/08/75411A
; Patent No. 6140072
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,411A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
```

```
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA: US 06/927,258
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-9/RFT/MTK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-475-411A-66

Query Match 72.6%; Score 45; DB 2; Length 357;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
Db 76 GPKGADGSGPG 85

RESULT 59
US-08-478-029A-66
; Sequence 66, Application US/08/78029A
; Patent No. 6184348
; GENERAL INFORMATION:
; APPLICANT: Ferrari, Franco A.
; APPLICANT: Cappello, Joseph
; TITLE OF INVENTION: Functional Recombinantly Prepared
; TITLE OF INVENTION: Synthetic Protein Polymer
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,029A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/609,716
; FILING DATE: 06-NOV-1990
; PRIOR APPLICATION NUMBER: US 07/269,429
; FILING DATE: 09-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/114,618
; FILING DATE: 29-OCT-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 06/927,258
; FILING DATE: 04-NOV-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-55186-8/RFT/MTK
; TELECOMMUNICATION INFORMATION:
```

; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-029A-66

Query Match 72.6%; Score 45; DB 2; Length 357;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
|:|||||
Db 76 GPKGADGSPG 85

RESULT 60

US-09-996-611D-4
; Sequence 4, Application US/09996611D
; Patent No. 6903200
; GENERAL INFORMATION:
; APPLICANT: Chou, Min-Yuan
; APPLICANT: Leu, Chyang-Yih
; TITLE OF INVENTION: No. 6903200el Human alpha 1 Chain Collagen
; FILE REFERENCE: 32350-176844
; CURRENT APPLICATION NUMBER: US/09/996,611D
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: Taiwan 89128027
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO. 4
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-611D-4

Query Match 72.6%; Score 45; DB 2; Length 509;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPGLL 12
|:|||||
Db 181 GKGAGGMPGLM 192

RESULT 61

US-09-219-849-48
; Sequence 48, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 595

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-48

Query Match 72.6%; Score 45; DB 2; Length 595;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
|:|||||
Db 570 GPKGADGSPG 579

RESULT 62

US-09-219-849-50
; Sequence 50, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 50
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-50

Query Match 72.6%; Score 45; DB 2; Length 595;
Best Local Similarity 80.0%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGSPG 10
|:|||||
Db 570 GPKGADGSPG 579

RESULT 63

US-09-219-849-49
; Sequence 49, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHEL D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849

```
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 49
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: amino acid sequence
US-09-219-849-49

Query Match 72.6%; Score 45; DB 2; Length 822;
Best Local Similarity 80.0%; Pred. No. 48;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPG 10
Db 570 GPKGADGSPG 579

RESULT 64
US-09-949-016-9992
; Sequence 9992, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9992
; LENGTH: 938
; TYPE: PRT
; ORGANISM: Human
; ORGANISM: Homo sapiens
US-09-949-016-9992

Query Match 72.6%; Score 45; DB 2; Length 938;
Best Local Similarity 72.7%; Pred. No. 55;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGL 11
Db 653 GKGAGVGQPGI 663

RESULT 65
US-09-996-611D-1
; Sequence 1, Application US/09996611D
; Patent No. 6903200
; GENERAL INFORMATION:
; APPLICANT: Chou, Min-Yuan
; APPLICANT: Leu, Chang-Yih
; TITLE OF INVENTION: NO. 6903200e1 Human alpha 1 Chain Collagen
; FILE REFERENCE: 32350-176844
; CURRENT APPLICATION NUMBER: US/09/996,611D
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: Taiwan 89128027
; PRIOR FILING DATE: 2000-12-27
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 954
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-996-611D-1

Query Match 72.6%; Score 45; DB 2; Length 954;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGGSPGLL 12
Db 626 GKGAGQMPGLM 637

RESULT 66
US-08-931-820-1
; Sequence 1, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type I
US-08-931-820-1

Query Match 72.6%; Score 45; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGGSPG 10
Db 588 GPKGADGSPG 597

RESULT 67
US-10-153-469A-16
; Sequence 16, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; APPLICANT: BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
```


ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-Oct-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-153-469A-16

Query Match 72.6%; Score 45; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSGPG 597

RESULT 68
US-10-153-469A-20
Sequence 20, Application US/10153469A
Patent No. 6927287
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BROKAW, JANE
BUECHTER, DOUGLAS
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-Oct-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-153-469A-20

Query Match 72.6%; Score 45; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSGPG 597

RESULT 69
US-10-104-889-16
Sequence 16, Application US/10104889
Patent No. 6958223
GENERAL INFORMATION:
APPLICANT: GRUSKIN, ELLIOT A.
BUECHTER, DOUGLAS
BROKAW, JANE
ZHANG, GUANGHUI
PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-Oct-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 1057 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-104-889-16

Query Match 72.6%; Score 45; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSGPG 597

RESULT 70
US-10-104-889-20
; Sequence 20, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-104-889-20
Query Match 72.6%; Score 45; DB 2; Length 1057;
Best Local Similarity 80.0%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GKGAGGSPG 10
Db 588 GPKGADGSPG 597
RESULT 71
US-10-153-469A-11
; Sequence 11, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY

COUNTRY: U.S.A.
ZIP: 11553
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1107 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-153-469A-11
Query Match 72.6%; Score 45; DB 2; Length 1107;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GKGAGGSPG 10
Db 588 GPKGADGSPG 597
RESULT 72
US-10-104-889-11
; Sequence 11, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
; TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DILWORTH & BARRESE
; STREET: 333 EARLE OVINGTON BOULEVARD
; CITY: UNIONDALE
; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/104,889
; FILING DATE: 22-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484

TELEFAX: (516) 228-8516
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1107 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-104-889-11.

Query Match 72.6%; Score 45; DB 2; Length 1107;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSPG 597

RESULT 73
US-10-153-469A-6
; Sequence 6, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BROKAW, JANE
; BUECHTER, DOUGLAS
; ZHANG, GUANGHUI
; PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,469A
FILING DATE: 22-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1169 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-153-469A-6

Query Match 72.6%; Score 45; DB 2; Length 1169;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSPG 597

RESULT 74
US-10-104-889-6
; Sequence 6, Application US/10104889
; Patent No. 6958223
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID
TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE
STATE: NY
COUNTRY: U.S.A.
ZIP: 11553

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/104,889
FILING DATE: 22-Mar-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/169,768
FILING DATE: 09-OCT-1998
ATTORNEY/AGENT INFORMATION:
NAME: STEEN, JEFFREY S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 228-8484
TELEFAX: (516) 228-8516

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1169 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-104-889-6

Query Match 72.6%; Score 45; DB 2; Length 1169;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
| | | | |
Db 588 GPKGADGSPG 597

RESULT 75
US-10-153-469A-8
; Sequence 8, Application US/10153469A
; Patent No. 6927287
; GENERAL INFORMATION:
; APPLICANT: GRUSKIN, ELLIOT A.
; BUECHTER, DOUGLAS
; BROKAW, JANE
; ZHANG, GUANGHUI
; PAOLELLA, DAVID

TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: DILWORTH & BARRESE
STREET: 333 EARLE OVINGTON BOULEVARD
CITY: UNIONDALE

```

; STATE: NY
; COUNTRY: U.S.A.
; ZIP: 11553
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153.469A
; FILING DATE: 22-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/169,768
; FILING DATE: 09-OCT-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEN, JEFFREY S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 228-8484
; TELEFAX: (516) 228-8516
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1171 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-469A-8

Query Match 72.6%; Score 45; DB 2; Length 1171;
Best Local Similarity 80.0%; Pred. No. 68;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSGPG 10
Db 588 GPGAGSGPG 597
| |||: |||

```

Search completed: March 11, 2006, 12:07:29
Job time : 42.6 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:06:37 ; Search time 140.4 Seconds
(without alignments)
35.712 Million cell updates/sec

Title: US-10-698-121A-2

Perfect score: 62

Sequence: 1 GEKAGSGPGLL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications_AA Main:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	62	100.0	12	5	US-10-698-121A-2
2	58	93.5	448	5	US-10-719-993-480
3	58	93.5	485	5	US-10-719-993-481
4	58	93.5	485	5	US-10-719-993-501
5	58	93.5	504	5	US-10-719-993-495
6	58	93.5	507	5	US-10-719-993-500
7	58	93.5	526	5	US-10-719-993-475
8	58	93.5	526	5	US-10-719-993-494
9	58	93.5	528	5	US-10-719-993-512
10	58	93.5	536	5	US-10-719-993-493
11	58	93.5	536	5	US-10-719-993-509
12	58	93.5	546	5	US-10-719-993-477
13	58	93.5	550	5	US-10-719-993-498
14	58	93.5	551	5	US-10-719-993-511
15	58	93.5	555	5	US-10-719-993-479
16	58	93.5	558	5	US-10-719-993-503
17	58	93.5	560	5	US-10-719-993-507
18	58	93.5	562	5	US-10-719-993-483
19	58	93.5	563	5	US-10-719-993-490
20	58	93.5	565	5	US-10-719-993-486
21	58	93.5	577	5	US-10-719-993-506
22	58	93.5	613	5	US-10-719-993-508
23	58	93.5	655	5	US-10-719-993-476
24	58	93.5	660	5	US-10-719-993-499
25	58	93.5	663	5	US-10-719-993-484
26	58	93.5	667	5	US-10-719-993-487
27	58	93.5	679	5	US-10-719-993-489

28	58	93.5	691	5	US-10-719-993-491	Sequence 491, App
29	58	93.5	693	5	US-10-719-993-502	Sequence 502, App
30	58	93.5	694	5	US-10-719-993-504	Sequence 504, App
31	58	93.5	698	5	US-10-719-993-497	Sequence 497, App
32	58	93.5	701	5	US-10-719-993-492	Sequence 492, App
33	58	93.5	703	5	US-10-719-993-488	Sequence 488, App
34	58	93.5	705	5	US-10-719-993-482	Sequence 482, App
35	58	93.5	706	5	US-10-719-993-496	Sequence 496, App
36	58	93.5	708	5	US-10-719-993-485	Sequence 510, App
37	58	93.5	708	5	US-10-719-993-510	Sequence 478, App
38	58	93.5	717	5	US-10-719-993-478	Sequence 505, App
39	58	93.5	720	5	US-10-719-993-505	Sequence 4225, App
40	52	83.9	1603	5	US-10-723-860-4225	Sequence 9, Appli
41	49	79.0	1712	3	US-09-961-403-9	Sequence 4, Appli
42	49	79.0	1712	3	US-10-648-813-4	Sequence 78, Appl
43	48	77.4	309	3	US-09-908-711-78	Sequence 21, Appl
44	48	77.4	1078	4	US-10-058-124-21	Sequence 226, App
45	48	77.4	1466	3	US-09-918-715-226	Sequence 68, Appl
46	48	77.4	1466	4	US-10-177-293-68	Sequence 33, Appl
47	48	77.4	1466	4	US-10-301-822-53	Sequence 72, Appl
48	48	77.4	1466	4	US-10-257-021-72	Sequence 4, Appli
49	48	77.4	1466	4	US-10-402-089-4	Sequence 6, Appli
50	48	77.4	1466	4	US-10-402-089-6	Sequence 12, Appl
51	48	77.4	1466	4	US-10-402-089-12	Sequence 4, Appli
52	48	77.4	1466	4	US-10-402-072A-4	Sequence 6, Appli
53	48	77.4	1466	4	US-10-402-072A-6	Sequence 12, Appl
54	48	77.4	1466	4	US-10-402-072A-12	Sequence 3, Appli
55	48	77.4	1466	4	US-10-357-851-3	Sequence 3, Appli
56	48	77.4	1466	4	US-10-358-024-3	Sequence 103, App
57	48	77.4	1466	4	US-10-734-564-103	Sequence 22, App
58	48	77.4	1466	4	US-10-474-794-226	Sequence 159, App
59	48	77.4	1466	5	US-10-852-335A-159	Sequence 226, App
60	48	77.4	1466	5	US-10-979-159-226	Sequence 451, App
61	48	77.4	1466	5	US-10-287-436A-451	Sequence 494, App
62	48	77.4	1466	5	US-10-287-436A-494	Sequence 1151, App
63	48	77.4	1466	5	US-10-287-436A-1151	Sequence 1189, App
64	48	77.4	1469	5	US-10-287-436A-1189	Sequence 45550, A
65	48	77.4	153	5	US-10-450-763-45550	Sequence 24, Appl
66	47	75.8	166	5	US-10-659-782A-24	Sequence 15, Appl
67	47	75.8	166	5	US-10-659-782A-25	Sequence 23, Appl
68	47	75.8	187	4	US-10-325-717-13	Sequence 32, Appl
69	47	75.8	187	4	US-10-325-717-32	Sequence 34, Appl
70	47	75.8	187	4	US-10-325-717-34	Sequence 50, Appl
71	47	75.8	187	4	US-10-325-717-50	Sequence 51, Appl
72	47	75.8	187	4	US-10-325-717-51	Sequence 52, Appl
73	47	75.8	187	4	US-10-325-717-52	Sequence 29, Appl
74	47	75.8	187	4	US-10-325-717-12	Sequence 30, Appl
75	47	75.8	193	4	US-10-325-717-29	Sequence 31, Appl
76	47	75.8	193	4	US-10-325-717-30	Sequence 47, Appl
77	47	75.8	193	4	US-10-325-717-31	Sequence 48, Appl
78	47	75.8	193	4	US-10-325-717-49	Sequence 49, Appl
79	47	75.8	193	4	US-10-325-717-48	Sequence 3, Appli
80	47	75.8	193	4	US-10-325-717-39	Sequence 20, Appl
81	47	75.8	203	4	US-10-325-717-20	Sequence 21, Appl
82	47	75.8	203	4	US-10-325-717-3	Sequence 22, Appl
83	47	75.8	203	4	US-10-325-717-21	Sequence 38, Appl
84	47	75.8	203	4	US-10-325-717-38	Sequence 39, Appl
85	47	75.8	203	4	US-10-325-717-39	Sequence 40, Appl
86	47	75.8	203	4	US-10-325-717-40	Sequence 2, Appli
87	47	75.8	226	4	US-10-325-717-2	Sequence 2, Appli
88	47	75.8	226	4	US-10-189-493-2	Sequence 2, Appli
89	47	75.8	231	4	US-10-189-493-2	Sequence 2, Appli
90	47	75.8	231	6	US-11-087-334-2	Sequence 14, Appl
91	47	75.8	233	4	US-10-379-747-14	Sequence 40, App
92	47	75.8	243	5	US-10-496-905-402	Sequence 6, Appli
93	47	75.8	243	5	US-09-776-976-6	Sequence 6, Appli
94	47	75.8	244	3	US-09-758-055-6	Sequence 48, Appl
95	47	75.8	244	3	US-09-909-547-6	Sequence 51, Appl
96	47	75.8	244	3	US-03-911-176B-48	Sequence 6, Appli
97	47	75.8	244	3	US-10-180-762-51	Sequence 4, Appli
98	47	75.8	244	4	US-10-231-814-6	Sequence 4, Appli
99	47	75.8	244	4	US-10-234-000-4	
100	47	75.8	244	4		

101	47	75.8	244	4	US-10-360-186-51	Sequence 51, Appl	174	45	72.6	251	4	US-10-232-175-21	Sequence 21, Appl
102	47	75.8	244	4	US-10-321-164-3	Sequence 3, Appl	175	45	72.6	251	4	US-10-232-175-31	Sequence 31, Appl
103	47	75.8	244	4	US-10-376-460-6	Sequence 6, Appl	176	45	72.6	251	6	US-11-139-377-21	Sequence 21, Appl
104	47	75.8	244	4	US-10-189-493-1	Sequence 1, Appl	177	45	72.6	251	6	US-11-139-377-31	Sequence 31, Appl
105	47	75.8	244	4	US-10-197-993-3	Sequence 3, Appl	178	45	72.6	416	6	US-10-232-175-25	Sequence 25, Appl
106	47	75.8	244	4	US-10-325-717-1	Sequence 1, Appl	179	45	72.6	416	6	US-11-139-377-25	Sequence 25, Appl
107	47	75.8	244	4	US-10-411-120-67	Sequence 67, Appl	180	45	72.6	429	3	US-09-764-864-1019	Sequence 1019, Ap
108	47	75.8	244	4	US-10-379-747-6	Sequence 6, Appl	181	45	72.6	483	5	US-10-758-846-88	Sequence 88, Ap
109	47	75.8	244	4	US-10-379-747-8	Sequence 8, Appl	182	45	72.6	498	5	US-10-901-816A-1	Sequence 1, Appl
110	47	75.8	244	4	US-10-379-747-10	Sequence 10, Appl	183	45	72.6	498	5	US-10-901-816A-2	Sequence 2, Appl
111	47	75.8	244	4	US-10-621-787-3	Sequence 3, Appl	184	45	72.6	500	4	US-10-232-175-22	Sequence 22, Appl
112	47	75.8	244	4	US-10-285-833-6	Sequence 6, Appl	185	45	72.6	500	6	US-11-139-377-22	Sequence 22, Appl
113	47	75.8	244	4	US-10-723-860-1048	Sequence 1048, Ap	186	45	72.6	510	4	US-10-232-175-26	Sequence 26, Appl
114	47	75.8	244	5	US-10-676-790-4	Sequence 4, Appl	187	45	72.6	510	6	US-11-139-377-26	Sequence 26, Appl
115	47	75.8	244	5	US-10-487-039-1	Sequence 1, Appl	188	45	72.6	544	5	US-10-658-989A-4	Sequence 4, Appl
116	47	75.8	244	5	US-10-775-180-681	Sequence 681, App	189	45	72.6	595	4	US-10-342-331-48	Sequence 48, Appl
117	47	75.8	244	5	US-10-775-180-682	Sequence 682, App	190	45	72.6	595	4	US-10-342-331-50	Sequence 50, Appl
118	47	75.8	244	5	US-10-775-180-683	Sequence 683, App	191	45	72.6	660	5	US-10-901-816A-3	Sequence 3, Appl
119	47	75.8	244	5	US-10-659-782A-22	Sequence 22, Appl	192	45	72.6	660	5	US-10-901-816A-4	Sequence 4, Appl
120	47	75.8	244	5	US-10-970-648-6	Sequence 6, Appl	193	45	72.6	662	4	US-10-232-175-33	Sequence 33, Appl
121	47	75.8	244	5	US-10-492-179-22	Sequence 22, Appl	194	45	72.6	662	6	US-11-139-377-33	Sequence 33, Appl
122	47	75.8	244	5	US-10-495-317-6	Sequence 6, Appl	195	45	72.6	723	6	US-11-097-143-13179	Sequence 13179, A
123	47	75.8	244	5	US-10-756-149-4970	Sequence 4970, Ap	196	45	72.6	822	4	US-10-342-331-49	Sequence 49, Appl
124	47	75.8	244	5	US-10-756-149-4970	Sequence 1785, Ap	197	45	72.6	1014	5	US-10-901-816A-5	Sequence 5, Appl
125	47	75.8	244	5	US-10-775-204-1785	Sequence 1785, Ap	198	45	72.6	1014	5	US-10-901-816A-6	Sequence 6, Appl
126	47	75.8	244	5	US-10-775-204-1785	Sequence 1787, Ap	199	45	72.6	1014	5	US-10-901-816A-7	Sequence 7, Appl
127	47	75.8	244	5	US-10-496-905-404	Sequence 404, App	200	45	72.6	1014	5	US-10-901-816A-8	Sequence 8, Appl
128	47	75.8	244	6	US-11-097-334-1	Sequence 1, Appl	201	45	72.6	1014	5	US-10-901-816A-9	Sequence 9, Appl
129	47	75.8	244	6	US-11-133-823-3	Sequence 3, Appl	202	45	72.6	1014	5	US-10-901-816A-10	Sequence 10, Appl
130	47	75.8	250	4	US-10-379-747-12	Sequence 12, Appl	203	45	72.6	1014	5	US-10-901-816A-11	Sequence 11, Appl
131	47	75.8	254	4	US-10-185-425-8	Sequence 8, Appl	204	45	72.6	1014	5	US-10-901-816A-13	Sequence 13, Appl
132	47	75.8	254	5	US-10-871-776-8	Sequence 8, Appl	205	45	72.6	1057	4	US-10-104-889-16	Sequence 16, Appl
133	47	75.8	836	5	US-10-775-180-608	Sequence 608, App	206	45	72.6	1057	4	US-10-104-889-20	Sequence 20, Appl
134	47	75.8	836	5	US-10-775-204-1619	Sequence 1619, Ap	207	45	72.6	1057	5	US-10-104-793-16	Sequence 16, Appl
135	47	75.8	1152	4	US-10-322-696-75	Sequence 75, Appl	208	45	72.6	1107	5	US-10-104-793-20	Sequence 20, Appl
136	47	75.8	1266	4	US-10-367-094-85	Sequence 85, Appl	209	45	72.6	1107	5	US-10-104-889-11	Sequence 11, Appl
137	47	75.8	1679	4	US-10-367-094-87	Sequence 87, Appl	210	45	72.6	1107	5	US-10-104-793-11	Sequence 11, Appl
138	47	75.8	1679	4	US-10-367-094-89	Sequence 89, Appl	211	45	72.6	1169	4	US-10-104-889-6	Sequence 6, Appl
139	47	75.8	1685	5	US-10-450-763-35198	Sequence 35198, A	212	45	72.6	1169	5	US-10-104-793-6	Sequence 6, Appl
140	47	75.8	1693	5	US-10-450-763-45978	Sequence 45978, A	213	45	72.6	1171	4	US-10-104-889-8	Sequence 8, Appl
141	46	74.2	492	4	US-10-639-286-11	Sequence 11, Appl	214	45	72.6	1171	5	US-10-104-793-8	Sequence 8, Appl
142	46	74.2	492	4	US-10-450-763-52407	Sequence 52407, A	215	45	72.6	1212	5	US-10-104-793-8	Sequence 8, Appl
143	46	74.2	518	5	US-10-233-885-44	Sequence 44, Appl	216	45	72.6	1212	5	US-10-331-496A-26	Sequence 26, Appl
144	46	74.2	714	4	US-10-231-581-44	Sequence 44, Appl	217	45	72.6	1341	4	US-10-852-335A-155	Sequence 155, App
145	46	74.2	714	4	US-10-231-581-44	Sequence 15, Appl	218	45	72.6	1341	4	US-10-124-18	Sequence 18, Appl
146	46	74.2	714	4	US-10-326-508A-15	Sequence 15, Appl	219	45	72.6	1388	4	US-10-058-124-18	Sequence 18, Appl
147	46	74.2	1014	4	US-10-194-441A-1	Sequence 1, Appl	220	45	72.6	1388	5	US-10-104-889-10	Sequence 10, Appl
148	46	74.2	1014	4	US-10-194-441A-48	Sequence 48, Appl	221	45	72.6	1461	4	US-10-468-091-26	Sequence 26, Appl
149	46	74.2	1014	4	US-10-639-286-10	Sequence 10, Appl	222	45	72.6	1461	4	US-10-468-091-25	Sequence 25, Appl
150	46	74.2	1418	4	US-10-058-124-20	Sequence 20, Appl	223	45	72.6	1464	3	US-09-918-715-261	Sequence 261, App
151	46	74.2	1418	4	US-10-468-091-5	Sequence 5, Appl	224	45	72.6	1464	4	US-10-060-036-159	Sequence 159, App
152	46	74.2	1487	4	US-10-468-091-6	Sequence 6, Appl	225	45	72.6	1464	4	US-10-171-311-36	Sequence 36, Appl
153	46	74.2	1487	5	US-10-756-149-4739	Sequence 4739, Ap	226	45	72.6	1464	4	US-10-216-705-21	Sequence 21, Appl
154	46	74.2	1629	5	US-10-287-436A-443	Sequence 443, App	227	45	72.6	1464	4	US-10-149-352-2	Sequence 2, Appl
155	46	74.2	1629	5	US-10-287-436A-443	Sequence 1444, Ap	228	45	72.6	1464	4	US-10-177-293-65	Sequence 65, Appl
156	46	74.2	1739	3	US-09-795-061-2	Sequence 2, Appl	229	45	72.6	1464	4	US-10-301-822-28	Sequence 28, Appl
157	46	74.2	1744	4	US-10-369-493-5832	Sequence 5832, Ap	230	45	72.6	1464	4	US-10-291-265-243	Sequence 243, App
158	46	74.2	1745	3	US-09-795-061-4	Sequence 4, Appl	231	45	72.6	1464	4	US-10-764-425-114	Sequence 114, App
159	46	74.2	1745	5	US-10-723-860-2660	Sequence 2660, Ap	232	45	72.6	1464	4	US-10-357-851-1	Sequence 1, Appl
160	46	74.2	1767	5	US-10-741-600-1476	Sequence 1476, Ap	233	45	72.6	1464	4	US-10-358-024-1	Sequence 79, Appl
161	46	74.2	1767	5	US-10-741-600-1476	Sequence 1480, Ap	234	45	72.6	1464	4	US-10-734-564-79	Sequence 150, Appl
162	46	74.2	1806	3	US-09-919-497-56	Sequence 56, Appl	235	45	72.6	1464	4	US-10-788-792-150	Sequence 261, App
163	46	74.2	1806	4	US-10-058-270A-122	Sequence 122, App	236	45	72.6	1464	5	US-10-852-335A-157	Sequence 157, App
164	46	74.2	1806	5	US-10-741-600-1478	Sequence 1478, Ap	237	45	72.6	1464	5	US-10-979-159-261	Sequence 491, App
165	46	74.2	1806	5	US-10-741-600-1479	Sequence 1479, Ap	238	45	72.6	1464	5	US-10-287-436A-491	Sequence 1187, Ap
166	46	74.2	1806	5	US-10-287-436A-498	Sequence 498, App	239	45	72.6	1669	3	US-10-287-436A-1187	Sequence 252, App
167	46	74.2	1806	5	US-10-287-436A-498	Sequence 1192, Ap	240	45	72.6	1669	4	US-09-918-715-252	Sequence 8, Appl
168	46	74.2	1818	5	US-10-741-600-1475	Sequence 1475, Ap	241	45	72.6	1669	4	US-10-372-683-8	Sequence 252, App
169	46	74.2	1818	5	US-10-741-600-1477	Sequence 1477, Ap	242	45	72.6	1669	4	US-10-474-794-252	Sequence 252, App
170	45	72.6	63	6	US-11-040-130-24	Sequence 20, Appl	243	45	72.6	1669	5	US-10-979-159-252	Sequence 252, App
171	45	72.6	63	6	US-11-007-053-20	Sequence 24, Appl	244	45	72.6	1669	5	US-10-979-159-252	Sequence 252, App
172	45	72.6	167	4	US-10-232-175-24	Sequence 24, Appl	245	45	72.6	1779	6	US-11-097-143-7413	Sequence 7413, Ap
173	45	72.6	167	6	US-11-139-377-24	Sequence 24, Appl	246	44	71.0	220	4	US-10-002-631C-264	Sequence 264, App

247	44	71.0	229	6	US-11-097-143-36510	Sequence 36510, A	320	43	69.4	520	4	US-10-173-706-332	Sequence 332, App
248	44	71.0	438	5	US-10-758-846-78	Sequence 78, App1	321	43	69.4	520	4	US-10-175-738-332	Sequence 332, App
249	44	71.0	459	4	US-10-331-496A-27	Sequence 27, App1	322	43	69.4	520	4	US-10-175-752-332	Sequence 332, App
250	44	71.0	459	4	US-10-372-683-30	Sequence 30, App1	323	43	69.4	520	4	US-10-176-482-332	Sequence 332, App
251	44	71.0	459	4	US-10-408-765A-2554	Sequence 2554, App	324	43	69.4	520	4	US-10-176-757-332	Sequence 332, App
252	44	71.0	459	5	US-10-758-846-28	Sequence 28, App1	325	43	69.4	520	4	US-10-176-913-332	Sequence 332, App
253	44	71.0	459	5	US-10-852-335A-156	Sequence 156, App	326	43	69.4	520	4	US-10-180-552-332	Sequence 332, App
254	44	71.0	476	5	US-10-758-846-91	Sequence 91, App1	327	43	69.4	520	4	US-10-180-557-332	Sequence 332, App
255	44	71.0	477	5	US-10-758-846-27	Sequence 27, App1	328	43	69.4	520	4	US-10-173-700-332	Sequence 332, App
256	44	71.0	616	5	US-10-450-763-49815	Sequence 49815, A	329	43	69.4	520	4	US-10-174-572-332	Sequence 332, App
257	44	71.0	680	4	US-10-177-293-59	Sequence 59, App1	330	43	69.4	520	4	US-10-174-579-332	Sequence 332, App
258	44	71.0	680	4	US-10-295-027-110	Sequence 110, App	331	43	69.4	520	4	US-10-174-582-332	Sequence 332, App
259	44	71.0	680	4	US-10-058-270A-80	Sequence 80, App1	332	43	69.4	520	4	US-10-174-588-332	Sequence 332, App
260	44	71.0	680	5	US-10-643-795A-118	Sequence 118, App	333	43	69.4	520	4	US-10-175-739-332	Sequence 332, App
261	44	71.0	680	5	US-10-948-518-118	Sequence 118, App	334	43	69.4	520	4	US-10-175-740-332	Sequence 332, App
262	44	71.0	680	5	US-10-936-626-83	Sequence 83, App1	335	43	69.4	520	4	US-10-175-743-332	Sequence 332, App
263	44	71.0	680	5	US-10-938-061-83	Sequence 83, App1	336	43	69.4	520	4	US-10-176-488-332	Sequence 332, App
264	44	71.0	698	5	US-10-732-923-1761	Sequence 1761, App	337	43	69.4	520	4	US-10-176-492-332	Sequence 332, App
265	44	71.0	840	4	US-10-367-094-82	Sequence 82, App1	338	43	69.4	520	4	US-10-176-747-332	Sequence 332, App
266	44	71.0	1051	5	US-10-450-763-50487	Sequence 50487, A	339	43	69.4	520	4	US-10-176-985-332	Sequence 332, App
267	43	69.4	234	3	US-09-895-674-1	Sequence 1, App1	340	43	69.4	520	4	US-10-176-987-332	Sequence 332, App
268	43	69.4	285	6	US-11-057-027-27	Sequence 27, App1	341	43	69.4	520	4	US-10-176-992-332	Sequence 332, App
269	43	69.4	335	5	US-10-820-155-128	Sequence 128, App	342	43	69.4	520	4	US-10-176-993-332	Sequence 332, App
270	43	69.4	399	5	US-10-858-412-239	Sequence 239, App	343	43	69.4	520	4	US-10-184-658-332	Sequence 332, App
271	43	69.4	403	3	US-09-925-302-689	Sequence 689, App	344	43	69.4	520	4	US-10-176-991-332	Sequence 332, App
272	43	69.4	403	3	US-09-925-302-689	Sequence 689, App	345	43	69.4	520	4	US-10-173-695-332	Sequence 332, App
273	43	69.4	407	5	US-10-741-600-1630	Sequence 1630, App	346	43	69.4	520	4	US-10-173-697-332	Sequence 332, App
274	43	69.4	433	5	US-10-858-412-240	Sequence 240, App	347	43	69.4	520	4	US-10-173-705-332	Sequence 332, App
275	43	69.4	481	5	US-10-758-846-93	Sequence 93, App1	348	43	69.4	520	4	US-10-174-576-332	Sequence 332, App
276	43	69.4	482	5	US-10-758-846-92	Sequence 92, App1	349	43	69.4	520	4	US-10-174-585-332	Sequence 332, App
277	43	69.4	493	5	US-10-758-846-89	Sequence 89, App1	350	43	69.4	520	4	US-10-174-586-332	Sequence 332, App
278	43	69.4	520	3	US-09-978-295A-614	Sequence 614, App	351	43	69.4	520	4	US-10-175-747-332	Sequence 332, App
279	43	69.4	520	3	US-09-978-697-614	Sequence 614, App	352	43	69.4	520	4	US-10-175-747-332	Sequence 332, App
280	43	69.4	520	3	US-09-978-132A-614	Sequence 614, App	353	43	69.4	520	4	US-10-176-481-332	Sequence 332, App
281	43	69.4	520	3	US-09-978-189-614	Sequence 614, App	354	43	69.4	520	4	US-10-176-485-332	Sequence 332, App
282	43	69.4	520	3	US-09-978-608A-614	Sequence 614, App	355	43	69.4	520	4	US-10-176-487-332	Sequence 332, App
283	43	69.4	520	3	US-09-978-189-614	Sequence 614, App	356	43	69.4	520	4	US-10-176-493-332	Sequence 332, App
284	43	69.4	520	3	US-09-978-585A-614	Sequence 614, App	357	43	69.4	520	4	US-10-176-756-332	Sequence 332, App
285	43	69.4	520	3	US-09-978-191A-614	Sequence 614, App	358	43	69.4	520	4	US-10-176-911-332	Sequence 332, App
286	43	69.4	520	3	US-09-978-403A-614	Sequence 614, App	359	43	69.4	520	4	US-10-176-919-332	Sequence 332, App
287	43	69.4	520	3	US-09-978-564A-614	Sequence 614, App	360	43	69.4	520	4	US-10-176-925-332	Sequence 332, App
288	43	69.4	520	3	US-09-999-833A-614	Sequence 614, App	361	43	69.4	520	4	US-10-176-978-332	Sequence 332, App
289	43	69.4	520	3	US-09-981-915A-614	Sequence 614, App	362	43	69.4	520	4	US-10-179-510-332	Sequence 332, App
290	43	69.4	520	3	US-09-978-824-614	Sequence 614, App	363	43	69.4	520	4	US-10-180-543-332	Sequence 332, App
291	43	69.4	520	3	US-09-918-585A-614	Sequence 614, App	364	43	69.4	520	4	US-10-180-544-332	Sequence 332, App
292	43	69.4	520	3	US-09-978-187B-614	Sequence 614, App	365	43	69.4	520	4	US-10-180-546-332	Sequence 332, App
293	43	69.4	520	3	US-09-999-834A-614	Sequence 614, App	366	43	69.4	520	4	US-10-180-547-332	Sequence 332, App
294	43	69.4	520	3	US-09-978-193A-614	Sequence 614, App	367	43	69.4	520	4	US-10-180-549-332	Sequence 332, App
295	43	69.4	520	3	US-09-999-830A-614	Sequence 614, App	368	43	69.4	520	4	US-10-180-555-332	Sequence 332, App
296	43	69.4	520	3	US-09-978-757A-614	Sequence 614, App	369	43	69.4	520	4	US-10-180-559-332	Sequence 332, App
297	43	69.4	520	3	US-09-978-187B-614	Sequence 614, App	370	43	69.4	520	4	US-10-181-000-332	Sequence 332, App
298	43	69.4	520	3	US-09-978-375A-614	Sequence 614, App	371	43	69.4	520	4	US-10-183-012-332	Sequence 332, App
299	43	69.4	520	3	US-09-978-298A-614	Sequence 614, App	372	43	69.4	520	4	US-10-183-012-332	Sequence 332, App
300	43	69.4	520	3	US-09-978-188A-614	Sequence 614, App	373	43	69.4	520	4	US-10-184-614-332	Sequence 332, App
301	43	69.4	520	3	US-09-978-188A-614	Sequence 614, App	374	43	69.4	520	4	US-10-184-635-332	Sequence 332, App
302	43	69.4	520	3	US-09-978-681A-614	Sequence 614, App	375	43	69.4	520	4	US-10-184-637-332	Sequence 332, App
303	43	69.4	520	3	US-09-978-194A-614	Sequence 614, App	376	43	69.4	520	4	US-10-184-646-332	Sequence 332, App
304	43	69.4	520	3	US-09-999-829A-614	Sequence 614, App	377	43	69.4	520	4	US-10-184-647-332	Sequence 332, App
305	43	69.4	520	3	US-09-978-299A-614	Sequence 614, App	378	43	69.4	520	4	US-10-184-652-332	Sequence 332, App
306	43	69.4	520	3	US-09-978-544A-614	Sequence 614, App	379	43	69.4	520	4	US-10-187-596-332	Sequence 332, App
307	43	69.4	520	3	US-09-978-665A-614	Sequence 614, App	380	43	69.4	520	4	US-10-187-594-332	Sequence 332, App
308	43	69.4	520	3	US-09-978-802A-614	Sequence 614, App	381	43	69.4	520	4	US-10-187-596-332	Sequence 332, App
309	43	69.4	520	3	US-09-999-831A-614	Sequence 614, App	382	43	69.4	520	4	US-10-187-885-332	Sequence 332, App
310	43	69.4	520	3	US-09-978-824A-614	Sequence 614, App	383	43	69.4	520	4	US-10-187-886-332	Sequence 332, App
311	43	69.4	520	4	US-10-052-586-332	Sequence 332, App	384	43	69.4	520	4	US-10-199-464-332	Sequence 332, App
312	43	69.4	520	4	US-10-174-590-332	Sequence 332, App	385	43	69.4	520	4	US-10-196-756-332	Sequence 332, App
313	43	69.4	520	4	US-10-176-758-332	Sequence 332, App	386	43	69.4	520	4	US-10-196-756-332	Sequence 332, App
314	43	69.4	520	4	US-10-175-737-332	Sequence 332, App	387	43	69.4	520	4	US-10-176-751-332	Sequence 332, App
315	43	69.4	520	4	US-10-174-581-332	Sequence 332, App	388	43	69.4	520	4	US-10-176-760-332	Sequence 332, App
316	43	69.4	520	4	US-10-176-483-332	Sequence 332, App	389	43	69.4	520	4	US-10-176-900-332	Sequence 332, App
317	43	69.4	520	4	US-10-176-749-332	Sequence 332, App	390	43	69.4	520	4	US-10-180-541-332	Sequence 332, App
318	43	69.4	520	4	US-10-176-914-332	Sequence 332, App	391	43	69.4	520	4	US-10-180-542-332	Sequence 332, App
319	43	69.4	520	4	US-10-176-915-332	Sequence 332, App	392	43	69.4	520	4	US-10-180-548-332	Sequence 332, App


```
US-10-719-993-480
; Sequence 480, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 480
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-993-480

Query Match      93.5%; Score 58; DB 5; Length 448;
Best Local Similarity 100.0%; Pred. No. 0.75;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      140 GEGGAEGSPGL 150

RESULT 3
US-10-719-993-481
; Sequence 481, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 481
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-993-481

Query Match      93.5%; Score 58; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      121 GEGGAEGSPGL 131

RESULT 4
US-10-719-993-501
; Sequence 501, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-993-501

Query Match      93.5%; Score 58; DB 5; Length 485;
Best Local Similarity 100.0%; Pred. No. 0.81;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      121 GEGGAEGSPGL 131

RESULT 5
US-10-719-993-495
; Sequence 495, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 495
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-993-495

Query Match      93.5%; Score 58; DB 5; Length 504;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      140 GEGGAEGSPGL 150

RESULT 6
US-10-719-993-500
; Sequence 500, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 500
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-719-993-500

Query Match      93.5%; Score 58; DB 5; Length 507;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      143 GEGGAEGSPGL 153

RESULT 7
US-10-719-993-475
; Sequence 475, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-475

Query Match          93.5%; Score 58; DB 5; Length 526;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      162 GKGAGSGPGL 172

RESULT 8
US-10-719-993-494
; Sequence 494, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 494
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-494

Query Match          93.5%; Score 58; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      172 GKGAGSGPGL 182

RESULT 11
US-10-719-993-509
; Sequence 509, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-509

Query Match          93.5%; Score 58; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      213 GKGAGSGPGL 223

RESULT 12
US-10-719-993-477
; Sequence 477, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 477
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-477

Query Match          93.5%; Score 58; DB 5; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      162 GKGAGSGPGL 172

RESULT 9
US-10-719-993-512
; Sequence 512, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 512
; LENGTH: 528
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-512

Query Match          93.5%; Score 58; DB 5; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      162 GKGAGSGPGL 172

RESULT 10
US-10-719-993-493
; Sequence 493, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 493
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-493

Query Match          93.5%; Score 58; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      172 GKGAGSGPGL 182

RESULT 11
US-10-719-993-509
; Sequence 509, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 509
; LENGTH: 536
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-509

Query Match          93.5%; Score 58; DB 5; Length 536;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      213 GKGAGSGPGL 223

RESULT 12
US-10-719-993-477
; Sequence 477, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 477
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-477

Query Match          93.5%; Score 58; DB 5; Length 528;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GKGAGSGPGL 11
Db      162 GKGAGSGPGL 172
```

```
; SEQ ID NO 477
; LENGTH: 546
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-477

Query Match      93.5%; Score 58; DB 5; Length 546;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      194 GEGGAEGSPGL 204

RESULT 13
US-10-719-993-498
; Sequence 498, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 498
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-498

Query Match      93.5%; Score 58; DB 5; Length 550;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      213 GEGGAEGSPGL 223

RESULT 14
US-10-719-993-511
; Sequence 511, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 511
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-511

Query Match      93.5%; Score 58; DB 5; Length 551;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      213 GEGGAEGSPGL 223

RESULT 15
US-10-719-993-479
```

```
; Sequence 479, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 479
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-479

Query Match      93.5%; Score 58; DB 5; Length 555;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      191 GEGGAEGSPGL 201

RESULT 16
US-10-719-993-503
; Sequence 503, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 503
; LENGTH: 558
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-503

Query Match      93.5%; Score 58; DB 5; Length 558;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
      |||||
Db      194 GEGGAEGSPGL 204

RESULT 17
US-10-719-993-507
; Sequence 507, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 507
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-507
```

```
Query Match      93.5%; Score 58; DB 5; Length 560;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      196 GEGGAGSPGL 206

RESULT 18
US-10-719-993-483
; Sequence 483, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 483
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-483

Query Match      93.5%; Score 58; DB 5; Length 562;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223

RESULT 19
US-10-719-993-490
; Sequence 490, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 490
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-490

Query Match      93.5%; Score 58; DB 5; Length 563;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223

RESULT 20
US-10-719-993-486
; Sequence 486, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 486
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-486

Query Match      93.5%; Score 58; DB 5; Length 565;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223
```

```
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 486
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-486

Query Match      93.5%; Score 58; DB 5; Length 565;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223

RESULT 21
US-10-719-993-506
; Sequence 506, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 506
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-506

Query Match      93.5%; Score 58; DB 5; Length 577;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223

RESULT 22
US-10-719-993-508
; Sequence 508, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 508
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-508

Query Match      93.5%; Score 58; DB 5; Length 613;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
Db      213 GEGGAGSPGL 223
```

Db 305 GEGGAEGSPGL 315

RESULT 23

US-10-719-993-476

; Sequence 476, Application US/10719993

; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 476

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-719-993-476

Query Match

Best Local Similarity 93.5%; Score 58; DB 5; Length 655;

Publication No. US20040265849A1

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11

|||||

Db 291 GEGGAEGSPGL 301

RESULT 24

US-10-719-993-499

; Sequence 499, Application US/10719993

; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 499

; LENGTH: 660

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-719-993-499

Query Match

Best Local Similarity 93.5%; Score 58; DB 5; Length 660;

Publication No. US20040265849A1

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11

|||||

Db 296 GEGGAEGSPGL 306

RESULT 25

US-10-719-993-484

; Sequence 484, Application US/10719993

; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 484

; LENGTH: 663

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-719-993-484

Query Match

Best Local Similarity 93.5%; Score 58; DB 5; Length 663;

Publication No. US20040265849A1

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11

|||||

Db 299 GEGGAEGSPGL 309

RESULT 26

US-10-719-993-487

; Sequence 487, Application US/10719993

; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 487

; LENGTH: 667

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-719-993-487

Query Match

Best Local Similarity 93.5%; Score 58; DB 5; Length 667;

Publication No. US20040265849A1

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11

|||||

Db 303 GEGGAEGSPGL 313

RESULT 27

US-10-719-993-489

; Sequence 489, Application US/10719993

; Publication No. US20040265849A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF

; FILE REFERENCE: CL001496

; CURRENT APPLICATION NUMBER: US/10/719,993

; CURRENT FILING DATE: 2003-11-24

; NUMBER OF SEQ ID NOS: 55342

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 489

; LENGTH: 679

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-719-993-489

Query Match

Best Local Similarity 93.5%; Score 58; DB 5; Length 679;

Publication No. US20040265849A1

Mismatches 0; Conservative 0; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11

|||||

Db 315 GEGGAEGSPGL 325

RESULT 28

US-10-719-993-491

; Sequence 491, Application US/10719993

```
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 491
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-491

Query Match      93.5%; Score 58; DB 5; Length 691;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      327 GEGGAEGSPGL 337

RESULT 29
US-10-719-993-502
; Sequence 502, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 502
; LENGTH: 693
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-502

Query Match      93.5%; Score 58; DB 5; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      356 GEGGAEGSPGL 366

RESULT 30
US-10-719-993-504
; Sequence 504, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 504
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-504

Query Match      93.5%; Score 58; DB 5; Length 694;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      356 GEGGAEGSPGL 366

RESULT 31
US-10-719-993-497
; Sequence 497, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 497
; LENGTH: 698
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-497

Query Match      93.5%; Score 58; DB 5; Length 698;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      356 GEGGAEGSPGL 366

RESULT 32
US-10-719-993-492
; Sequence 492, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 492
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-492

Query Match      93.5%; Score 58; DB 5; Length 701;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      334 GEGGAEGSPGL 344

RESULT 33
US-10-719-993-488
; Sequence 488, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT FILING DATE: 2003-11-24
; CURRENT APPLICATION NUMBER: US/10/719,993
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-488

Query Match      93.5%; Score 58; DB 5; Length 694;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAEGSPGL 11
Db      337 GEGGAEGSPGL 347
```


; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 488
; LENGTH: 703
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-488

Query Match 93.5%; Score 58; DB 5; Length 703;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGGSPGL 11
|||||
Db 339 GEGGAGGSPGL 349

RESULT 34

US-10-719-993-482
; Sequence 482, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 482
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-482

Query Match 93.5%; Score 58; DB 5; Length 705;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGGSPGL 11
|||||
Db 356 GEGGAGGSPGL 366

RESULT 35

US-10-719-993-496
; Sequence 496, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 496
; LENGTH: 706
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-496

Query Match 93.5%; Score 58; DB 5; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGGSPGL 11
|||||
Db 356 GEGGAGGSPGL 366

RESULT 36

US-10-719-993-485
; Sequence 485, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 485
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-485

Query Match 93.5%; Score 58; DB 5; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGGSPGL 11
|||||
Db 356 GEGGAGGSPGL 366

RESULT 37

US-10-719-993-510
; Sequence 510, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 510
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-510

Query Match 93.5%; Score 58; DB 5; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGGSPGL 11
|||||
Db 344 GEGGAGGSPGL 354

RESULT 38

US-10-719-993-478
; Sequence 478, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 478
; LENGTH: 717

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-478

Query Match      93.5%; Score 58; DB 5; Length 717;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSGPGL 11
Db      353 GEGGAGSGPGL 363

RESULT 39
US-10-719-993-505
; Sequence 505, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 5342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 505
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-505

Query Match      93.5%; Score 58; DB 5; Length 720;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSGPGL 11
Db      356 GEGGAGSGPGL 366

RESULT 40
US-10-723-860-4225
; Sequence 4225, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4225
; LENGTH: 1603
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-4225

Query Match      83.9%; Score 52; DB 5; Length 1603;
Best Local Similarity 81.8%; Pred. No. 24;
Matches 9; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEGGAGSGPGL 11
Db      817 GEGGAGSGPGL 827

US-10-698-121a-2.rapbm
```

```
RESULT 41
US-09-961-403-9
; Sequence 9, Application US/09961403
; Publication No. US20030077589A1
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-9

Query Match      79.0%; Score 49; DB 3; Length 1712;
Best Local Similarity 72.7%; Pred. No. 76;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGGAGSGPGL 11
Db      276 GEGGAGSGPGL 286

RESULT 42
US-10-648-813-4
; Sequence 4, Application US/10648813
; Publication No. US20050048063A1
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Essler, Markus
; APPLICANT: Brown, Darren M.
; TITLE OF INVENTION: Collagen-Binding Molecules That
; TITLE OF INVENTION: Selectively Home To Tumor Vasculature and Methods of Using
; TITLE OF INVENTION: Same
; FILE REFERENCE: 66821-233
; CURRENT APPLICATION NUMBER: US/10/648,813
; CURRENT FILING DATE: 2003-08-25
; PRIOR APPLICATION NUMBER: US 10/233,153
; PRIOR FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-813-4

Query Match      79.0%; Score 49; DB 5; Length 1712;
Best Local Similarity 72.7%; Pred. No. 76;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGGAGSGPGL 11
Db      276 GEGGAGSGPGL 286

RESULT 43
US-09-908-711-78
; Sequence 78, Application US/09908711
; Patent No. US20020045230A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
```

Mon Mar 13 10:53:09 2006

FILE REFERENCE: PA128
CURRENT APPLICATION NUMBER: US/09/908,711
CURRENT FILING DATE: 2001-07-20
PRIOR APPLICATION NUMBER: US01/01360
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,867
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01344
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,892
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01345
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,888
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01329
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,905
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01354
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,891
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01339
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,869
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01340
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,874
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01334
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,853
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01349
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,902
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01239
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,870
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01348
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,882
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01347
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,896
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01307
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,864
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01341
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,856
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01336
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/764,868
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: US01/01312
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04

PRIOR APPLICATION NUMBER: 60/209,467
PRIOR FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 167
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 78
LENGTH: 309
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (4)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-78
Query Match 77.4%; Score 48; DB 3; Length 309;
Best Local Similarity 80.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 GKGAGGSPG 10
||:|||||
Db 13 GERSEGGSPG 22
||:|||||
RESULT 44
US-10-058-124-21
Sequence 21, Application US/10058124
Publication No. US20030119058A1
GENERAL INFORMATION:
APPLICANT: Qvist, Per
Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
in Body Fluids, A Test Kit and Means for Carrying Out the
Method and Use of the Method to Diagnose the Presence of
Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/058,124
FILING DATE: 29-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/570,573
FILING DATE: 2002-MAY-12
APPLICATION NUMBER: 08/187,319
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1078 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein


```
Db      1170 GERSESPG 1179
||:|:|||||
US-10-257-021-72
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 6
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-6
Query Match      77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEKAGSPG 10
||:|:|||||
Db      1170 GERSESPG 1179

RESULT 51
US-10-402-089-12
; Sequence 12, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-089-12
Query Match      77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEKAGSPG 10
||:|:|||||
Db      1171 GERSESPG 1180

RESULT 52
US-10-402-072A-4
; Sequence 4, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A

Db      1170 GERSESPG 1179
||:|:|||||
US-10-257-021-72
; Publication No. US20030211498A1
; GENERAL INFORMATION:
; APPLICANT: Morin, Patrice J.
; APPLICANT: Sherman-Baust, Cheryl A.
; APPLICANT: Pizer, Ellen S.
; APPLICANT: Hough, Colleen D.
; TITLE OF INVENTION: TUMOR MARKERS IN OVARIAN CANCER
; FILE REFERENCE: 14014.0369U2
; CURRENT APPLICATION NUMBER: US/10/257,021
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: PCT/US01/10947
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,336
; PRIOR FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-257-021-72
Query Match      77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEKAGSPG 10
||:|:|||||
Db      1170 GERSESPG 1179

RESULT 49
US-10-402-089-4
; Sequence 4, Application US/10402089
; Publication No. US20040005663A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: PORCINE COLLAGENS AND GELATINS
; FILE REFERENCE: FP0402.3 CON
; CURRENT APPLICATION NUMBER: US/10/402,089
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 4
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-089-4
Query Match      77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GEKAGSPG 10
||:|:|||||
Db      1170 GERSESPG 1179

RESULT 50
US-10-402-089-6
; Sequence 6, Application US/10402089
```

; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Bos Taurus
US-10-402-072A-4

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 53

US-10-402-072A-6
; Sequence 6, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: EP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa
US-10-402-072A-6

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 54

US-10-402-072A-12
; Sequence 12, Application US/10402072A
; Publication No. US20040018592A1
; GENERAL INFORMATION:
; APPLICANT: Bell, Marcum P.
; APPLICANT: Neff, Thomas B.
; APPLICANT: Polarek, James W.
; APPLICANT: Seeley, Todd W.
; TITLE OF INVENTION: BOVINE COLLAGENS AND GELATINS
; FILE REFERENCE: EP0402.2 CON
; CURRENT APPLICATION NUMBER: US/10/402,072A
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: US 09/709,700
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Sus scrofa

US-10-402-072A-12

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|||||
Db 1171 GERSESGSPG 1180

RESULT 55

US-10-357-851-3
; Sequence 3, Application US/10357851
; Publication No. US20040151731A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; TITLE OF INVENTION: Method and Compositions Involving
; TITLE OF INVENTION: T-Lymphocyte Reactivity with Collagen in Aortic Tissue of
; TITLE OF INVENTION: Abdominal Aortic Aneurysm Patients
; FILE REFERENCE: 13376US
; CURRENT APPLICATION NUMBER: US/10/357,851
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-357-851-3

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 56

US-10-358-024-3
; Sequence 3, Application US/10358024
; Publication No. US20040151732A1
; GENERAL INFORMATION:
; APPLICANT: Jicha, Douglas L.
; APPLICANT: Pelsue, Stephen
; TITLE OF INVENTION: Methods and Compositions Involving Blood
; TITLE OF INVENTION: T-Lymphocytes Reactivity with Collagen
; FILE REFERENCE: 13436US
; CURRENT APPLICATION NUMBER: US/10/358,024
; CURRENT FILING DATE: 2003-02-04
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-358-024-3

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 57

US-10-734-564-103
; Sequence 103, Application US/10734564
; Publication No. US20040157278A1

GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using T1MP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-103

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 1170 GERGEGSPG 1179

RESULT 58
US-10-474-794-226
; Sequence 226, Application US/10474794
; Publication No. US20040213793A1
; GENERAL INFORMATION:
; APPLICANT: Carbon-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: US/10/474,794
; CURRENT FILING DATE: 2003-10-14
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-474-794-226

Query Match 77.4%; Score 48; DB 4; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 1170 GERGEGSPG 1179

RESULT 59
US-10-852-335A-159
; Sequence 159, Application US/10852335A
; Publication No. US20050112129A1
; GENERAL INFORMATION:
; APPLICANT: HEIDI S. PHILLIPS
; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; TITLE OF INVENTION: Treatment of Tumors of Glial Origin
; FILE REFERENCE: P5103R1-US
; CURRENT APPLICATION NUMBER: US/10/852,335A
; CURRENT FILING DATE: 2004-05-24
; PRIOR APPLICATION NUMBER: US 60/548,299
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/473,238
; PRIOR FILING DATE: 2003-05-23
; NUMBER OF SEQ ID NOS: 190

; SEQ ID NO 159
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-852-335A-159

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 1170 GERGEGSPG 1179

RESULT 60
US-10-979-159-226
; Sequence 226, Application US/10979159
; Publication No. US20050142138A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT APPLICATION NUMBER: US/10/979,159
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US/09/918,715
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 226
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-979-159-226

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAEGSPG 10
Db 1170 GERGEGSPG 1179

RESULT 61
US-10-287-436A-451
; Sequence 451, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
; APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
; TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS
; FILE REFERENCE: 10872.514696
; CURRENT APPLICATION NUMBER: US/10/287,436A
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/336,220
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 1446
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 451
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-287-436A-451

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 62

US-10-287-436A-494
; Sequence 494, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
; APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
; TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 10872.514696
; CURRENT APPLICATION NUMBER: US/10/287,436A
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/336,220
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 1446
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 494
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-287-436A-494

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 63

US-10-287-436A-1151
; Sequence 1151, Application US/10287436A
; Publication No. US20050202421A1
; GENERAL INFORMATION:
; APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
; TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 10872.514696
; CURRENT APPLICATION NUMBER: US/10/287,436A
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/336,220
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 1446
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1151
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-287-436A-1151

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 64

US-10-287-436A-1189
; Sequence 1189, Application US/10287436A
; Publication No. US20050202421A1

; GENERAL INFORMATION:
; APPLICANT: CHILDREN'S HOSPITAL MEDICAL CENTER
; TITLE OF INVENTION: METHOD FOR DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 10872.514696
; CURRENT APPLICATION NUMBER: US/10/287,436A
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/336,220
; PRIOR FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 1446
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1189
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-287-436A-1189

Query Match 77.4%; Score 48; DB 5; Length 1466;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSPG 10
||:|||||
Db 1170 GERSESGSPG 1179

RESULT 65

US-10-450-763-45550
; Sequence 45550, Application US/10450763
; Publication No. US20050196754A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 790CIP3/US
; CURRENT APPLICATION NUMBER: US/10/450,763
; CURRENT FILING DATE: 2003-06-11
; PRIOR APPLICATION NUMBER: PCT/US01/08631
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 45550
; LENGTH: 1469
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (1258)..(1306)
; OTHER INFORMATION: 352 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL domain identified by
; OTHER INFORMATION: eMATRIX, accession number DM01418A, p-value=1.000e-40, raw score
; OTHER INFORMATION: of 20.83
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (85)..(1199)
; OTHER INFORMATION: Collagen triple helix repeat (20 copies) domain identified by
; OTHER INFORMATION: Pfam, accession name Collagen, E-value=6.8e-282, Pfam score of
; OTHER INFORMATION: 949.9
US-10-450-763-45550

Query Match 77.4%; Score 48; DB 5; Length 1469;
Best Local Similarity 80.0%; Pred. No. 94;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGGAGSPG 10
||:|||||
Db 1173 GERSESGSPG 1182

RESULT 66

US-10-659-782A-24

```

; Sequence 24, Application US/10659782A
; Publication No. US20050059015A1
; GENERAL INFORMATION:
; APPLICANT: Mintz, Liat
; TITLE OF INVENTION: Compositions, Reagents and Kits for and Methods of Diagnosing,
; FILE REFERENCE: 28238
; CURRENT APPLICATION NUMBER: US/10/659,782A
; CURRENT FILING DATE: 2003-09-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 24
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-659-782A-24

Query Match 75.8%; Score 47; DB 5; Length 153;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPGGLL 12
Db 63 GEGGAGSPGGLI 74

RESULT 67
US-10-659-782A-25
; Sequence 25, Application US/10659782A
; Publication No. US20050059015A1
; GENERAL INFORMATION:
; APPLICANT: Mintz, Liat
; TITLE OF INVENTION: Compositions, Reagents and Kits for and Methods of Diagnosing,
; FILE REFERENCE: 28238
; CURRENT APPLICATION NUMBER: US/10/659,782A
; CURRENT FILING DATE: 2003-09-11
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 25
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-659-782A-25

Query Match 75.8%; Score 47; DB 5; Length 166;
Best Local Similarity 66.7%; Pred. No. 15;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPGGLL 12
Db 63 GEGGAGSPGGLI 74

RESULT 68
US-10-325-717-13
; Sequence 13, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct T121C-apM1 (58-244)
US-10-325-717-32

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPGGLL 12
Db 6 GEGGAGSPGGLI 17

RESULT 70
US-10-325-717-33
; Sequence 33, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03

```

```

; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: apM1 (58-244)
US-10-325-717-13

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPGGLL 12
Db 6 GEGGAGSPGGLI 17

RESULT 69
US-10-325-717-32
; Sequence 32, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct T121C-apM1 (58-244)
US-10-325-717-32

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPGGLL 12
Db 6 GEGGAGSPGGLI 17

RESULT 70
US-10-325-717-33
; Sequence 33, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03

```

Db 6 GEKGEKDPGLI 17

RESULT 72

US-10-325-717-50

Sequence 50, Application US/10325717

Publication No. US20030176328A1

GENERAL INFORMATION:

APPLICANT: Rasmussen, Poul Baad

APPLICANT: Andersen, Kim Vilbour

APPLICANT: Pedersen, Anders Hjelholt

APPLICANT: Schambye, Hans Thalsgaard

APPLICANT: Halkier, Torben

APPLICANT: Bognes, Are

TITLE OF INVENTION: Adiponectin Fragments and Conjugates

FILE REFERENCE: 0251us610

CURRENT APPLICATION NUMBER: US/10/325,717

CURRENT FILING DATE: 2002-12-20

PRIOR APPLICATION NUMBER: US 60/412,169

PRIOR FILING DATE: 2002-09-20

PRIOR APPLICATION NUMBER: US 60/394,117

PRIOR FILING DATE: 2002-07-03

PRIOR APPLICATION NUMBER: US 60/375,492

PRIOR FILING DATE: 2002-04-25

PRIOR APPLICATION NUMBER: US 60/343,482

PRIOR FILING DATE: 2001-12-21

NUMBER OF SEQ ID NOS: 115

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 33

LENGTH: 187

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic construct S146C-apM1 (58-244)

US-10-325-717-33

Query Match 75.8%; Score 47; DB 4; Length 187;

Best Local Similarity 66.7%; Pred. No. 17;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGLL 12

|||||:||||:

Db 6 GEKGEKDPGLI 17

RESULT 71

US-10-325-717-34

Sequence 34, Application US/10325717

Publication No. US20030176328A1

GENERAL INFORMATION:

APPLICANT: Rasmussen, Poul Baad

APPLICANT: Andersen, Kim Vilbour

APPLICANT: Pedersen, Anders Hjelholt

APPLICANT: Schambye, Hans Thalsgaard

APPLICANT: Halkier, Torben

APPLICANT: Bognes, Are

TITLE OF INVENTION: Adiponectin Fragments and Conjugates

FILE REFERENCE: 0251us610

CURRENT APPLICATION NUMBER: US/10/325,717

CURRENT FILING DATE: 2002-12-20

PRIOR APPLICATION NUMBER: US 60/412,169

PRIOR FILING DATE: 2002-09-20

PRIOR APPLICATION NUMBER: US 60/394,117

PRIOR FILING DATE: 2002-07-03

PRIOR APPLICATION NUMBER: US 60/375,492

PRIOR FILING DATE: 2002-04-25

PRIOR APPLICATION NUMBER: US 60/343,482

PRIOR FILING DATE: 2001-12-21

NUMBER OF SEQ ID NOS: 115

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 34

LENGTH: 187

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Synthetic construct T243C-apM1 (58-244)

US-10-325-717-34

Query Match 75.8%; Score 47; DB 4; Length 187;

Best Local Similarity 66.7%; Pred. No. 17;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGLL 12

|||||:||||:

Query Match 75.8%; Score 47; DB 4; Length 187;

Best Local Similarity 66.7%; Pred. No. 17;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAEGSPGLL 12

|||||:||||:

Db 6 GEKGEKDPGLI 17

RESULT 73

US-10-325-717-51

Sequence 51, Application US/10325717

Publication No. US20030176328A1

GENERAL INFORMATION:

APPLICANT: Rasmussen, Poul Baad

APPLICANT: Andersen, Kim Vilbour

APPLICANT: Pedersen, Anders Hjelholt

APPLICANT: Schambye, Hans Thalsgaard

APPLICANT: Halkier, Torben

APPLICANT: Bognes, Are

TITLE OF INVENTION: Adiponectin Fragments and Conjugates

FILE REFERENCE: 0251us610

CURRENT APPLICATION NUMBER: US/10/325,717

CURRENT FILING DATE: 2002-12-20

PRIOR APPLICATION NUMBER: US 60/412,169

PRIOR FILING DATE: 2002-09-20

PRIOR APPLICATION NUMBER: US 60/394,117

PRIOR FILING DATE: 2002-07-03

PRIOR APPLICATION NUMBER: US 60/375,492

PRIOR FILING DATE: 2002-04-25

PRIOR APPLICATION NUMBER: US 60/343,482

PRIOR FILING DATE: 2001-12-21

NUMBER OF SEQ ID NOS: 115

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 51

LENGTH: 187

TYPE: PRT

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N141C-apM1 (58-244)
US-10-325-717-51

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 74
US-10-325-717-52
; Sequence 52, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N228C-apM1 (58-244)
US-10-325-717-52

Query Match 75.8%; Score 47; DB 4; Length 193;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 12 GEGGAGSPGGL 23

Search completed: March 11, 2006, 12:11:31
Job time : 143.4 secs

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N141C-apM1 (58-244)
US-10-325-717-51

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 74
US-10-325-717-52
; Sequence 52, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N228C-apM1 (58-244)
US-10-325-717-52

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 75
US-10-325-717-12
; Sequence 12, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N141C-apM1 (58-244)
US-10-325-717-51

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 74
US-10-325-717-52
; Sequence 52, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N228C-apM1 (58-244)
US-10-325-717-52

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 75
US-10-325-717-12
; Sequence 12, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20

; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N141C-apM1 (58-244)
US-10-325-717-51

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 74
US-10-325-717-52
; Sequence 52, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 60/394,117
; PRIOR FILING DATE: 2002-07-03
; PRIOR APPLICATION NUMBER: US 60/375,492
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US 60/343,482
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic construct N228C-apM1 (58-244)
US-10-325-717-52

Query Match 75.8%; Score 47; DB 4; Length 187;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGGAGSPGGL 12
Db 6 GEGGAGSPGGL 17

RESULT 75
US-10-325-717-12
; Sequence 12, Application US/10325717
; Publication No. US20030176328A1
; GENERAL INFORMATION:
; APPLICANT: Rasmussen, Poul Baad
; APPLICANT: Andersen, Kim Vilbour
; APPLICANT: Pedersen, Anders Hjelholt
; APPLICANT: Schambye, Hans Thalsgaard
; APPLICANT: Halkier, Torben
; APPLICANT: Bognes, Are
; TITLE OF INVENTION: Adiponectin Fragments and Conjugates
; FILE REFERENCE: 0251us610
; CURRENT APPLICATION NUMBER: US/10/325,717
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/412,169
; PRIOR FILING DATE: 2002-09-20

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2006, 12:07:47 ; Search time 16.8 Seconds
(without alignments)
19.882 Million cell updates/sec

Title: US-10-698-121A-2

Perfect score: 62

Sequence: 1 GKGAGCSPGLL 12

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 161667 seqs, 27934885 residues

Total number of hits satisfying chosen parameters: 161667

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 500 summaries

Database : Published Applications AA New:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	77.4	1166	6 US-10-821-234-964	Sequence 964, App
2	48	77.4	1466	7 US-11-186-284-33	Sequence 33, Appl
3	47	75.8	244	6 US-10-514-040-4	Sequence 4, Appl
4	47	75.8	244	6 US-10-296-865-6	Sequence 6, Appl
5	47	75.8	244	7 US-11-157-947-1	Sequence 1, Appl
6	47	75.8	244	7 US-11-256-802-3	Sequence 3, Appl
7	47	75.8	244	7 US-11-258-647-4	Sequence 4, Appl
8	46	74.2	1081	7 US-11-051-720-1372	Sequence 1372, Ap
9	46	74.2	1736	7 US-11-124-368A-329	Sequence 329, App
10	46	74.2	1767	6 US-10-995-561-911	Sequence 911, App
11	46	74.2	1767	6 US-10-995-561-914	Sequence 914, App
12	46	74.2	1806	6 US-10-995-561-912	Sequence 912, App
13	46	74.2	1806	6 US-10-995-561-915	Sequence 915, App
14	46	74.2	1806	7 US-11-051-720-1446	Sequence 1446, Ap
15	46	74.2	1806	7 US-11-051-720-1447	Sequence 1447, Ap
16	46	74.2	1818	6 US-10-995-561-910	Sequence 910, App
17	46	74.2	1818	6 US-10-995-561-913	Sequence 913, App
18	45	72.6	1464	6 US-10-501-035-331	Sequence 331, App
19	45	72.6	1464	7 US-11-000-463-243	Sequence 243, App
20	45	72.6	1464	7 US-11-186-284-28	Sequence 28, Appl
21	45	72.6	1464	7 US-11-021-603-2	Sequence 2, Appl
22	43	69.4	520	6 US-10-821-234-1096	Sequence 1096, Ap
23	43	69.4	520	6 US-10-995-561-532	Sequence 532, App
24	43	69.4	744	7 US-11-186-284-37	Sequence 37, Appl
25	43	69.4	1496	7 US-11-186-284-35	Sequence 35, Appl

26	42	67.7	186	7 US-11-181-091-6	Sequence 6, Appl
27	42	67.7	287	7 US-11-174-150-30	Sequence 30, Appl
28	42	67.7	484	7 US-11-181-091-7	Sequence 7, Appl
29	42	67.7	489	7 US-11-181-091-8	Sequence 8, Appl
30	42	67.7	822	6 US-10-330-773-700	Sequence 700, App
31	42	67.7	1366	6 US-10-821-234-1431	Sequence 1431, Ap
32	42	67.7	1366	7 US-11-186-284-31	Sequence 31, Appl
33	42	67.7	1874	6 US-10-821-234-1182	Sequence 1182, Ap
34	42	67.7	2551	6 US-11-052-554A-368	Sequence 368, App
35	41	66.1	66	6 US-10-514-057-1	Sequence 1, Appl
36	41	66.1	243	6 US-10-131-826A-362	Sequence 362, App
37	41	66.1	243	6 US-10-973-115B-362	Sequence 362, App
38	41	66.1	247	6 US-10-514-040-2	Sequence 2, Appl
39	41	66.1	247	6 US-10-236-865-2	Sequence 2, Appl
40	41	66.1	247	6 US-10-236-865-4	Sequence 4, Appl
41	41	66.1	334	6 US-10-514-057-6	Sequence 6, Appl
42	41	66.1	828	6 US-10-995-561-983	Sequence 983, App
43	41	66.1	918	6 US-10-995-561-981	Sequence 981, App
44	41	66.1	1019	6 US-10-995-561-982	Sequence 982, App
45	41	66.1	1532	6 US-10-821-234-914	Sequence 914, App
46	40	64.5	177	7 US-11-245-689-17	Sequence 17, Appl
47	40	64.5	180	7 US-11-245-689-18	Sequence 18, Appl
48	40	64.5	186	7 US-11-245-689-15	Sequence 15, Appl
49	40	64.5	288	7 US-11-135-855-30	Sequence 30, Appl
50	40	64.5	303	7 US-11-135-855-31	Sequence 31, Appl
51	40	64.5	303	7 US-11-258-647-2	Sequence 2, Appl
52	40	64.5	313	7 US-11-245-689-44	Sequence 44, App
53	40	64.5	326	6 US-10-055-877-289	Sequence 289, App
54	40	64.5	326	6 US-10-055-877-297	Sequence 297, App
55	40	64.5	629	6 US-10-821-234-1528	Sequence 1528, Ap
56	40	64.5	744	7 US-11-186-284-39	Sequence 39, Appl
57	40	64.5	749	7 US-11-052-554A-148	Sequence 148, App
58	39	62.9	75	7 US-11-245-689-19	Sequence 19, Appl
59	39	62.9	245	7 US-11-256-802-4	Sequence 4, Appl
60	39	62.9	258	7 US-11-051-720-1334	Sequence 1334, Ap
61	39	62.9	258	7 US-11-051-720-1434	Sequence 1434, Ap
62	39	62.9	259	7 US-11-122-524-2	Sequence 2, Appl
63	39	62.9	278	7 US-11-113-424-43	Sequence 43, Appl
64	39	62.9	285	7 US-11-258-647-5	Sequence 5, Appl
65	39	62.9	289	7 US-11-258-647-15	Sequence 15, Appl
66	39	62.9	571	7 US-11-072-512-3814	Sequence 3814, Ap
67	39	62.9	1028	7 US-11-169-041-180	Sequence 180, App
68	39	62.9	1516	6 US-10-220-824-8	Sequence 8, Appl
69	38	61.3	117	7 US-11-096-070-34	Sequence 34, Appl
70	38	61.3	117	7 US-11-245-689-32	Sequence 32, Appl
71	38	61.3	199	7 US-11-113-424-44	Sequence 44, Appl
72	38	61.3	234	7 US-11-245-689-36	Sequence 36, Appl
73	38	61.3	237	7 US-11-245-689-37	Sequence 37, Appl
74	38	61.3	237	7 US-11-245-689-41	Sequence 41, Appl
75	38	61.3	237	7 US-11-245-689-43	Sequence 43, Appl
76	38	61.3	280	5 US-09-978-360A-809	Sequence 809, App
77	38	61.3	280	6 US-10-821-234-1300	Sequence 1300, Ap
78	38	61.3	281	6 US-10-131-826A-372	Sequence 372, App
79	38	61.3	281	6 US-10-973-115B-372	Sequence 372, App
80	38	61.3	281	7 US-11-258-647-3	Sequence 3, Appl
81	38	61.3	284	7 US-11-170-268-40	Sequence 40, Appl
82	38	61.3	317	7 US-11-024-959-335	Sequence 335, App
83	38	61.3	341	7 US-11-170-268-28	Sequence 28, Appl
84	38	61.3	341	7 US-11-170-268-36	Sequence 36, Appl
85	38	61.3	470	7 US-11-072-512-3177	Sequence 3177, Ap
86	38	61.3	481	7 US-11-186-284-63	Sequence 63, Appl
87	38	61.3	531	7 US-11-096-070-4	Sequence 4, Appl
88	38	61.3	549	7 US-11-096-070-2	Sequence 2, Appl
89	38	61.3	549	7 US-11-096-070-6	Sequence 6, Appl
90	38	61.3	549	7 US-11-096-070-8	Sequence 8, Appl
91	38	61.3	615	7 US-11-052-554A-152	Sequence 152, App
92	38	61.3	729	7 US-11-051-720-1373	Sequence 1373, Ap
93	38	61.3	738	7 US-11-051-720-1374	Sequence 1374, Ap
94	38	61.3	1519	7 US-11-182-016-5	Sequence 5, Appl
95	37	59.7	19	6 US-10-503-595-124	Sequence 124, App
96	37	59.7	123	7 US-11-245-689-6	Sequence 6, Appl
97	37	59.7	126	7 US-11-245-689-4	Sequence 4, Appl
98	37	59.7	129	7 US-11-245-689-27	Sequence 27, Appl

99	37	59.7	132	7	US-11-245-689-23	Sequence 23, Appl	172	35	56.5	248	7	US-11-150-887-14	Sequence 14, Appl
100	37	59.7	135	7	US-11-245-689-11	Sequence 11, Appl	173	35	56.5	248	7	US-11-241-035-28	Sequence 28, Appl
101	37	59.7	135	7	US-11-245-689-28	Sequence 28, Appl	174	35	56.5	251	7	US-11-096-568A-18119	Sequence 18119, A
102	37	59.7	138	7	US-11-245-689-2	Sequence 2, Appl	175	35	56.5	252	7	US-11-170-268-26	Sequence 26, Appl
103	37	59.7	138	7	US-11-245-689-16	Sequence 16, Appl	176	35	56.5	264	7	US-11-096-568A-16150	Sequence 16150, A
104	37	59.7	141	7	US-11-245-689-25	Sequence 25, Appl	177	35	56.5	288	6	US-10-131-826A-316	Sequence 316, App
105	37	59.7	144	7	US-11-245-689-13	Sequence 13, Appl	178	35	56.5	288	6	US-10-973-115B-316	Sequence 316, App
106	37	59.7	144	7	US-11-245-689-1	Sequence 1, Appl	179	35	56.5	299	7	US-11-177-506-37	Sequence 37, Appl
107	37	59.7	150	7	US-11-245-689-10	Sequence 10, Appl	180	35	56.5	309	7	US-11-170-268-24	Sequence 24, Appl
108	37	59.7	150	7	US-11-245-689-10	Sequence 10, Appl	181	35	56.5	313	7	US-11-177-506-36	Sequence 36, Appl
109	37	59.7	156	7	US-11-096-568A-11972	Sequence 11972, A	182	35	56.5	333	7	US-11-096-568A-16148	Sequence 16148, A
110	37	59.7	165	7	US-11-245-689-24	Sequence 24, Appl	183	35	56.5	386	7	US-11-072-512-2777	Sequence 2777, Ap
111	37	59.7	165	7	US-11-245-689-20	Sequence 20, Appl	184	35	56.5	406	7	US-11-072-512-3291	Sequence 3291, Ap
112	37	59.7	187	6	US-10-055-877-296	Sequence 296, App	185	35	56.5	462	7	US-11-124-367A-497	Sequence 497, App
113	37	59.7	191	6	US-10-055-877-87	Sequence 87, Appl	186	35	56.5	538	7	US-11-124-368A-311	Sequence 311, App
114	37	59.7	191	6	US-10-453-372-916	Sequence 916, App	187	35	56.5	538	7	US-11-124-368A-312	Sequence 312, App
115	37	59.7	203	6	US-10-055-877-99	Sequence 99, Appl	188	35	56.5	538	7	US-11-124-368A-313	Sequence 313, App
116	37	59.7	266	6	US-10-055-877-95	Sequence 95, Appl	189	35	56.5	546	7	US-11-124-367A-498	Sequence 498, App
117	37	59.7	266	6	US-10-453-372-926	Sequence 926, App	190	35	56.5	546	7	US-11-124-367A-499	Sequence 499, App
118	37	59.7	283	6	US-10-055-877-91	Sequence 91, Appl	191	35	56.5	1306	7	US-11-052-554A-139	Sequence 139, App
119	37	59.7	283	6	US-10-453-372-922	Sequence 922, App	192	35	56.5	3093	7	US-11-186-284-26	Sequence 26, Appl
120	37	59.7	285	7	US-11-072-512-3028	Sequence 3028, Ap	193	34	54.8	76	7	US-11-105-708-4	Sequence 4, Appl
121	37	59.7	296	6	US-10-453-372-920	Sequence 920, App	194	34	54.8	76	7	US-11-096-568A-12579	Sequence 12579, A
122	37	59.7	299	6	US-10-055-877-97	Sequence 97, Appl	195	34	54.8	80	7	US-11-245-689-29	Sequence 29, Appl
123	37	59.7	304	6	US-10-055-877-83	Sequence 83, Appl	196	34	54.8	129	6	US-10-967-527A-13	Sequence 13, Appl
124	37	59.7	309	7	US-11-096-568A-18960	Sequence 18960, A	197	34	54.8	161	6	US-10-943-015-2	Sequence 2, Appl
125	37	59.7	316	6	US-10-453-372-924	Sequence 924, App	198	34	54.8	231	6	US-10-454-437-194	Sequence 194, App
126	37	59.7	318	6	US-10-802-796-727	Sequence 727, App	199	34	54.8	246	7	US-11-072-512-3694	Sequence 3694, Ap
127	37	59.7	319	6	US-10-055-877-93	Sequence 93, Appl	200	34	54.8	246	7	US-11-067-121-4	Sequence 4, Appl
128	37	59.7	319	6	US-10-055-877-288	Sequence 88, App	201	34	54.8	246	7	US-11-256-802-12	Sequence 12, Appl
129	37	59.7	319	6	US-10-055-877-294	Sequence 294, App	202	34	54.8	249	7	US-11-096-568A-2390	Sequence 2390, Ap
130	37	59.7	319	6	US-10-453-372-912	Sequence 912, App	203	34	54.8	249	7	US-11-096-568A-2393	Sequence 2393, Ap
131	37	59.7	319	6	US-10-453-372-932	Sequence 932, App	204	34	54.8	261	6	US-10-512-184-35	Sequence 35, Appl
132	37	59.7	326	6	US-10-055-877-286	Sequence 286, App	205	34	54.8	267	6	US-10-454-437-192	Sequence 192, App
133	37	59.7	326	6	US-10-055-877-287	Sequence 287, App	206	34	54.8	267	6	US-11-055-822-358	Sequence 358, App
134	37	59.7	326	6	US-10-055-877-293	Sequence 293, App	207	34	54.8	295	6	US-10-995-561-999	Sequence 999, App
135	37	59.7	326	6	US-10-055-877-295	Sequence 295, App	208	34	54.8	295	7	US-11-193-205-1	Sequence 1, Appl
136	37	59.7	326	6	US-10-453-372-998	Sequence 998, App	209	34	54.8	313	6	US-10-512-184-72	Sequence 72, Appl
137	37	59.7	326	6	US-10-453-372-930	Sequence 930, App	210	34	54.8	358	6	US-10-995-561-885	Sequence 885, App
138	37	59.7	334	6	US-10-802-796-728	Sequence 728, App	211	34	54.8	358	6	US-10-995-561-888	Sequence 889, App
139	37	59.7	441	7	US-11-100-640-30	Sequence 30, Appl	212	34	54.8	358	6	US-10-995-561-889	Sequence 4, Appl
140	37	59.7	783	7	US-11-052-554A-157	Sequence 157, App	213	34	54.8	373	6	US-10-515-547-4	Sequence 4, Appl
141	37	59.7	914	7	US-11-052-554A-160	Sequence 160, App	214	34	54.8	388	6	US-10-995-561-887	Sequence 887, App
142	36	58.1	114	7	US-11-245-689-21	Sequence 21, Appl	215	34	54.8	397	7	US-11-087-099-9323	Sequence 9323, Ap
143	36	58.1	120	7	US-11-096-070-35	Sequence 35, Appl	216	34	54.8	400	6	US-10-689-742-74	Sequence 74, Appl
144	36	58.1	253	6	US-10-821-234-1438	Sequence 1438, Ap	217	34	54.8	438	7	US-11-072-512-3166	Sequence 3166, App
145	36	58.1	260	6	US-10-485-517-354	Sequence 354, App	218	34	54.8	451	6	US-10-995-561-886	Sequence 886, App
146	36	58.1	261	6	US-10-485-517-150	Sequence 150, App	219	34	54.8	531	7	US-11-060-914-4	Sequence 4, Appl
147	36	58.1	350	7	US-11-096-568A-19022	Sequence 19022, A	220	34	54.8	546	7	US-11-143-980-38	Sequence 38, Appl
148	36	58.1	406	6	US-10-131-826A-82	Sequence 82, Appl	221	34	54.8	561	6	US-10-943-015-4	Sequence 4, Appl
149	36	58.1	406	6	US-11-096-070-82	Sequence 82, Appl	222	34	54.8	574	6	US-10-507-275-7	Sequence 7, Appl
150	36	58.1	427	7	US-11-096-070-10	Sequence 10, Appl	223	34	54.8	574	6	US-10-770-726-50	Sequence 50, Appl
151	36	58.1	442	7	US-11-096-568A-19021	Sequence 19021, A	224	34	54.8	581	7	US-11-045-802-30	Sequence 30, Appl
152	36	58.1	455	7	US-11-096-568A-19020	Sequence 19020, A	225	34	54.8	584	7	US-11-045-802-31	Sequence 31, Appl
153	36	58.1	551	7	US-11-096-070-12	Sequence 12, Appl	226	34	54.8	618	6	US-10-821-234-1481	Sequence 1481, Ap
154	36	58.1	711	6	US-10-517-939-4	Sequence 4, Appl	227	34	54.8	741	7	US-11-052-554A-161	Sequence 161, App
155	36	58.1	837	7	US-11-052-554A-159	Sequence 159, App	228	34	54.8	763	6	US-10-873-528-63	Sequence 63, Appl
156	36	58.1	1329	7	US-11-052-554A-136	Sequence 136, App	229	34	54.8	778	7	US-11-052-554A-144	Sequence 144, App
157	36	58.1	1329	7	US-11-087-099-882	Sequence 882, App	230	34	54.8	801	7	US-11-052-554A-165	Sequence 165, App
158	36	58.1	1408	7	US-11-087-099-8482	Sequence 8482, Ap	231	34	54.8	820	7	US-11-165-819-1	Sequence 1, Appl
159	36	58.1	2647	6	US-10-821-234-1303	Sequence 1303, Ap	232	34	54.8	915	6	US-10-821-234-1514	Sequence 1514, Ap
160	35	56.5	33	7	US-11-150-883-27	Sequence 27, Appl	233	34	54.8	915	6	US-10-995-561-1003	Sequence 1003, Ap
161	35	56.5	33	7	US-11-150-887-20	Sequence 20, Appl	234	34	54.8	917	6	US-10-995-561-1000	Sequence 1000, Ap
162	35	56.5	53	7	US-11-150-883-26	Sequence 26, Appl	235	34	54.8	929	7	US-11-087-099-6648	Sequence 6648, Ap
163	35	56.5	53	7	US-11-150-887-19	Sequence 19, Appl	236	34	54.8	929	7	US-11-087-099-7898	Sequence 7898, Ap
164	35	56.5	90	7	US-11-245-689-22	Sequence 22, Appl	237	34	54.8	940	6	US-10-995-561-1004	Sequence 1004, Ap
165	35	56.5	123	6	US-10-980-388-72	Sequence 72, Appl	238	34	54.8	969	6	US-10-995-561-1001	Sequence 1001, Ap
166	35	56.5	174	6	US-10-980-388-88	Sequence 88, Appl	239	34	54.8	971	6	US-10-995-561-998	Sequence 998, App
167	35	56.5	244	7	US-11-096-568A-18501	Sequence 18501, A	240	34	54.8	994	6	US-10-995-561-997	Sequence 997, App
168	35	56.5	246	7	US-11-067-121-13	Sequence 13, Appl	241	34	54.8	1122	6	US-10-821-234-1657	Sequence 1657, Ap
169	35	56.5	246	7	US-11-170-268-14	Sequence 14, Appl	242	34	54.8	1132	6	US-10-964-313-4	Sequence 4, Appl
170	35	56.5	246	7	US-11-256-802-2	Sequence 2, Appl	243	34	54.8	1660	7	US-11-052-554A-137	Sequence 137, App
171	35	56.5	248	7	US-11-150-883-21	Sequence 21, Appl	244	34	54.8	2392	6	US-10-330-773-907	Sequence 907, App

245	34	54.8	3375	7	US-11-044-111-23	Sequence 23, Appl	318	32	51.6	395	7	US-11-072-512-3413	Sequence 3413, Ap
246	33.5	54.0	162	7	US-11-072-512-2111	Sequence 1111, Ap	319	32	51.6	401	7	US-11-087-099-4694	Sequence 4694, Ap
247	33.5	54.0	164	7	US-11-087-227-18	Sequence 218, Appl	320	32	51.6	403	7	US-11-096-568A-23495	Sequence 23495, A
248	33.5	54.0	164	7	US-11-087-227-20	Sequence 20, Appl	321	32	51.6	415	7	US-11-096-568A-23494	Sequence 23494, A
249	33.5	54.0	272	7	US-11-240-769-114	Sequence 114, App	322	32	51.6	434	7	US-11-052-554A-167	Sequence 167, App
250	33	53.2	15	6	US-10-938-890-491	Sequence 491, App	323	32	51.6	467	7	US-11-096-568A-23493	Sequence 23493, A
251	33	53.2	79	7	US-11-245-689-9	Sequence 9, Appli	324	32	51.6	475	6	US-10-793-626-2076	Sequence 2076, Ap
252	33	53.2	102	7	US-11-245-689-8	Sequence 8, Appli	325	32	51.6	482	6	US-10-878-568A-139	Sequence 139, App
253	33	53.2	111	7	US-11-245-689-38	Sequence 38, Appl	326	32	51.6	484	7	US-11-126-313-24	Sequence 24, Appl
254	33	53.2	111	7	US-11-245-689-39	Sequence 39, Appl	327	32	51.6	494	7	US-11-087-099-3900	Sequence 3900, Ap
255	33	53.2	114	7	US-11-245-689-7	Sequence 7, Appli	328	32	51.6	525	6	US-10-793-626-1292	Sequence 1292, Ap
256	33	53.2	114	7	US-11-245-689-14	Sequence 14, Appl	329	32	51.6	541	7	US-11-087-099-7458	Sequence 7458, Ap
257	33	53.2	118	7	US-11-245-689-12	Sequence 12, Appl	330	32	51.6	542	7	US-11-074-176-30	Sequence 30, Appl
258	33	53.2	174	7	US-11-245-689-33	Sequence 33, Appl	331	32	51.6	558	6	US-10-714-995-46	Sequence 46, Appl
259	33	53.2	174	7	US-11-087-099-7376	Sequence 7376, Ap	332	32	51.6	576	7	US-11-052-554A-164	Sequence 164, App
260	33	53.2	179	7	US-11-197-133A-62	Sequence 62, Appl	333	32	51.6	585	7	US-11-241-347-15	Sequence 15, Appl
261	33	53.2	194	7	US-11-072-512-3133	Sequence 3133, Ap	334	32	51.6	602	7	US-11-072-512-3324	Sequence 3324, Ap
262	33	53.2	196	7	US-11-087-099-10837	Sequence 10837, A	335	32	51.6	608	7	US-11-226-701-8	Sequence 8, Appli
263	33	53.2	222	7	US-11-245-689-40	Sequence 40, Appl	336	32	51.6	608	7	US-11-241-347-9	Sequence 9, Appli
264	33	53.2	234	7	US-11-096-568A-9983	Sequence 9983, Ap	337	32	51.6	711	6	US-10-330-773-70	Sequence 70, Appl
265	33	53.2	243	6	US-11-063-703-122	Sequence 122, App	338	32	51.6	719	6	US-10-793-626-1548	Sequence 1548, Ap
266	33	53.2	243	7	US-11-102-240-122	Sequence 122, App	339	32	51.6	732	6	US-10-467-657-5888	Sequence 5888, Ap
267	33	53.2	243	7	US-11-080-991-14	Sequence 14, Appl	340	32	51.6	754	6	US-10-330-773-72	Sequence 72, Appl
268	33	53.2	244	7	US-11-096-568A-9982	Sequence 9982, Ap	341	32	51.6	772	6	US-10-330-773-67	Sequence 67, Appl
269	33	53.2	278	7	US-11-186-284-77	Sequence 77, Appl	342	32	51.6	853	7	US-11-052-554A-149	Sequence 149, App
270	33	53.2	319	6	US-11-055-877-290	Sequence 290, App	343	32	51.6	861	7	US-11-186-641A-2	Sequence 2, Appli
271	33	53.2	364	7	US-11-169-041-216	Sequence 216, App	344	32	51.6	905	7	US-11-072-512-2728	Sequence 2728, Ap
272	33	53.2	402	6	US-10-649-591-18	Sequence 18, Appl	345	32	51.6	924	6	US-10-857-780-20	Sequence 20, Appl
273	33	53.2	429	6	US-10-878-568A-17	Sequence 17, Appl	346	32	51.6	924	6	US-10-493-909-67	Sequence 67, Appl
274	33	53.2	485	7	US-11-096-568A-11820	Sequence 11820, A	347	32	51.6	924	6	US-11-107-028-26	Sequence 26, Appl
275	33	53.2	639	7	US-11-087-099-3621	Sequence 3621, Ap	348	32	51.6	981	7	US-11-087-099-3267	Sequence 3267, Ap
276	33	53.2	650	5	US-09-995-493-170	Sequence 170, App	349	32	51.6	1011	7	US-11-069-642-111	Sequence 111, App
277	33	53.2	737	6	US-10-501-035-254	Sequence 254, App	350	32	51.6	1036	7	US-11-072-512-2812	Sequence 2812, Ap
278	33	53.2	756	7	US-11-072-512-2505	Sequence 2505, App	351	32	51.6	1065	7	US-11-201-606-16	Sequence 16, Appl
279	33	53.2	923	6	US-10-793-626-3108	Sequence 3108, Ap	352	32	51.6	1336	6	US-10-912-971-10	Sequence 10, Appl
280	33	53.2	923	7	US-11-052-554A-147	Sequence 147, App	353	32	51.6	1538	7	US-11-052-554A-146	Sequence 146, App
281	33	53.2	1971	6	US-11-179-624-1	Sequence 1, Appli	354	32	51.6	1783	7	US-11-126-313-38	Sequence 38, Appl
282	33	53.2	1971	6	US-10-821-234-1262	Sequence 1262, Ap	355	32	51.6	1901	7	US-11-052-554A-135	Sequence 135, App
283	33	53.2	3069	7	US-11-235-732-2	Sequence 2, Appli	356	32	51.6	3132	7	US-11-087-099-1245	Sequence 1245, Ap
284	33	53.2	3095	7	US-11-235-732-4	Sequence 4, Appli	357	32	51.6	3132	7	US-11-108-172-1116	Sequence 1116, Ap
285	33	53.2	3105	7	US-11-235-732-7	Sequence 7, Appli	358	31.5	50.8	143	6	US-10-485-788A-733	Sequence 733, App
286	33	53.2	3105	7	US-11-235-732-7	Sequence 7, Appli	359	31.5	50.8	143	6	US-11-083-076-103	Sequence 103, App
287	33	53.2	3105	7	US-11-235-732-7	Sequence 7, Appli	360	31.5	50.8	445	7	US-11-087-099-8602	Sequence 8602, Ap
288	32.5	52.4	3467	7	US-10-453-372-1112	Sequence 1112, Ap	361	31	50.0	11	7	US-11-105-708-3	Sequence 3, Appli
289	32.5	52.4	357	7	US-11-096-568A-18376	Sequence 18376, A	362	31	50.0	11	7	US-11-105-708-8	Sequence 8, Appli
290	32.5	52.4	161	7	US-11-096-568A-18375	Sequence 18375, A	363	31	50.0	19	6	US-10-503-575-128	Sequence 128, App
291	32.5	52.4	267	7	US-11-096-568A-27043	Sequence 27043, A	364	31	50.0	19	6	US-10-503-575-132	Sequence 132, App
292	32.5	52.4	271	7	US-11-096-568A-27045	Sequence 27045, A	365	31	50.0	20	6	US-10-623-155-410	Sequence 410, App
293	32.5	52.4	383	7	US-11-096-568A-22981	Sequence 22981, A	366	31	50.0	20	6	US-10-623-155-521	Sequence 521, App
294	32.5	52.4	391	7	US-11-096-568A-22980	Sequence 22980, A	367	31	50.0	70	6	US-10-623-155-471	Sequence 471, App
295	32.5	52.4	401	7	US-11-096-568A-22979	Sequence 22979, A	368	31	50.0	93	7	US-11-096-568A-927	Sequence 927, App
296	32.5	52.4	495	7	US-11-087-099-6970	Sequence 6970, Ap	369	31	50.0	108	7	US-11-096-568A-926	Sequence 926, App
297	32	51.6	42	7	US-11-245-689-5	Sequence 5, Appli	370	31	50.0	113	7	US-11-072-512-3273	Sequence 3273, Ap
298	32	51.6	102	7	US-11-096-568A-24201	Sequence 24201, A	371	31	50.0	113	7	US-11-072-512-3273	Sequence 3273, Ap
299	32	51.6	118	7	US-11-096-568A-15267	Sequence 15267, A	372	31	50.0	114	7	US-11-096-568A-925	Sequence 925, App
300	32	51.6	121	5	US-09-978-360A-494	Sequence 494, App	373	31	50.0	117	7	US-11-087-099-5641	Sequence 5641, Ap
301	32	51.6	159	7	US-11-245-689-30	Sequence 30, Appl	374	31	50.0	128	7	US-11-096-568A-20859	Sequence 20859, A
302	32	51.6	183	7	US-11-245-689-34	Sequence 34, Appl	375	31	50.0	132	7	US-11-096-568A-885	Sequence 885, App
303	32	51.6	198	7	US-11-245-689-42	Sequence 42, Appl	376	31	50.0	138	6	US-10-821-234-1172	Sequence 1172, Ap
304	32	51.6	210	7	US-11-052-554A-328	Sequence 328, App	377	31	50.0	148	7	US-11-096-568A-20684	Sequence 20684, A
305	32	51.6	216	7	US-11-245-689-35	Sequence 35, Appl	378	31	50.0	151	7	US-11-072-512-3804	Sequence 3804, Ap
306	32	51.6	227	7	US-11-096-568A-25049	Sequence 25049, A	379	31	50.0	179	5	US-09-978-360A-566	Sequence 566, App
307	32	51.6	244	7	US-11-090-878-66	Sequence 66, Appl	380	31	50.0	185	7	US-11-072-512-1972	Sequence 1972, A
308	32	51.6	256	7	US-11-096-568A-25048	Sequence 25048, A	381	31	50.0	196	7	US-11-096-568A-10176	Sequence 10176, A
309	32	51.6	257	7	US-11-096-568A-24246	Sequence 24246, A	382	31	50.0	219	7	US-11-087-099-8268	Sequence 8268, Ap
310	32	51.6	293	6	US-11-096-568A-11808	Sequence 11808, A	383	31	50.0	252	7	US-11-096-568A-25298	Sequence 25298, A
311	32	51.6	304	6	US-10-793-626-1472	Sequence 1472, Ap	384	31	50.0	255	7	US-11-115-086-11	Sequence 11, Appl
312	32	51.6	308	6	US-10-055-877-85	Sequence 85, Appl	385	31	50.0	257	7	US-11-096-568A-5557	Sequence 5557, Ap
313	32	51.6	308	6	US-10-453-372-914	Sequence 914, App	386	31	50.0	257	7	US-11-096-568A-34230	Sequence 34230, A
314	32	51.6	313	7	US-11-096-568A-11807	Sequence 11807, A	387	31	50.0	262	7	US-11-087-099-9170	Sequence 9170, Ap
315	32	51.6	324	7	US-11-096-568A-11806	Sequence 11806, A	388	31	50.0	266	7	US-11-096-568A-23193	Sequence 23193, A
316	32	51.6	350	7	US-11-087-099-12163	Sequence 12163, A	389	31	50.0	270	7	US-11-096-568A-16837	Sequence 16837, A
317	32	51.6	362	7	US-11-072-512-2732	Sequence 2732, Ap	390	31	50.0	278	6	US-10-793-626-2344	Sequence 2344, Ap
			366	7	US-11-096-568A-11751	Sequence 11751, A				285	7	US-11-096-568A-6540	Sequence 6540, Ap

391	31	50.0	299	7	US-11-096-568A-10731	Sequence 10731, A	464	31	50.0	579	6	US-10-623-155-480	Sequence 480, App
392	31	50.0	302	6	US-10-453-372-780	Sequence 780, App	465	31	50.0	579	6	US-10-623-155-484	Sequence 484, App
393	31	50.0	302	6	US-10-453-372-782	Sequence 782, App	466	31	50.0	579	6	US-10-501-035-255	Sequence 255, App
394	31	50.0	302	6	US-10-453-372-788	Sequence 788, App	467	31	50.0	579	7	US-11-090-617-684	Sequence 684, App
395	31	50.0	304	6	US-10-453-372-792	Sequence 792, App	468	31	50.0	580	7	US-11-087-099-1273	Sequence 1273, App
396	31	50.0	304	6	US-11-096-568A-6539	Sequence 6539, App	469	31	50.0	580	7	US-11-087-099-6034	Sequence 6034, App
397	31	50.0	304	6	US-11-096-568A-20393	Sequence 20393, A	470	31	50.0	580	7	US-11-087-099-6200	Sequence 6200, App
398	31	50.0	308	6	US-10-453-372-778	Sequence 778, App	471	31	50.0	580	7	US-11-087-099-9615	Sequence 9615, App
399	31	50.0	308	6	US-10-453-372-778	Sequence 778, App	472	31	50.0	581	7	US-11-087-099-4940	Sequence 4940, App
400	31	50.0	318	6	US-11-080-029-21	Sequence 21, Appl	473	31	50.0	581	7	US-11-087-099-10168	Sequence 10168, A
401	31	50.0	329	6	US-10-821-234-884	Sequence 884, App	474	31	50.0	582	7	US-11-087-099-1734	Sequence 1734, App
402	31	50.0	329	6	US-11-125-402-28	Sequence 28, Appl	475	31	50.0	585	7	US-11-096-568A-11064	Sequence 11064, A
403	31	50.0	331	7	US-11-087-099-4126	Sequence 4126, App	476	31	50.0	586	6	US-10-623-155-427	Sequence 427, App
404	31	50.0	335	7	US-11-096-568A-5556	Sequence 5556, App	477	31	50.0	589	6	US-10-623-155-486	Sequence 486, App
405	31	50.0	336	7	US-11-096-568A-20392	Sequence 20392, A	478	31	50.0	599	6	US-11-096-568A-2151	Sequence 2151, App
406	31	50.0	340	7	US-11-096-568A-5535	Sequence 5535, App	479	31	50.0	606	7	US-11-052-554A-163	Sequence 163, App
407	31	50.0	344	7	US-11-060-029-15	Sequence 15, Appl	480	31	50.0	607	7	US-11-096-568A-11063	Sequence 11063, A
408	31	50.0	345	7	US-11-096-568A-10730	Sequence 10730, A	481	31	50.0	618	7	US-11-052-554A-150	Sequence 150, App
409	31	50.0	346	7	US-11-060-029-19	Sequence 19, Appl	482	31	50.0	628	7	US-11-080-991-108	Sequence 108, App
410	31	50.0	349	7	US-11-087-099-10937	Sequence 10937, A	483	31	50.0	639	7	US-11-074-176-222	Sequence 222, App
411	31	50.0	359	7	US-11-096-568A-9000	Sequence 9000, App	484	31	50.0	639	6	US-10-821-234-1016	Sequence 1016, App
412	31	50.0	359	7	US-11-096-568A-34229	Sequence 34229, A	485	31	50.0	652	6	US-11-052-554A-158	Sequence 158, App
413	31	50.0	360	7	US-11-096-568A-16836	Sequence 16836, A	486	31	50.0	694	7	US-10-793-626-2922	Sequence 2922, App
414	31	50.0	361	7	US-11-052-554A-169	Sequence 169, App	487	31	50.0	702	7	US-11-096-568A-14646	Sequence 14646, A
415	31	50.0	364	7	US-11-096-568A-34228	Sequence 34228, A	488	31	50.0	715	6	US-10-131-826A-116	Sequence 116, App
416	31	50.0	365	7	US-11-096-568A-17761	Sequence 17761, A	489	31	50.0	715	6	US-10-793-626-570	Sequence 570, App
417	31	50.0	382	7	US-11-096-568A-16835	Sequence 16835, A	490	31	50.0	715	6	US-10-973-115B-116	Sequence 116, App
418	31	50.0	388	6	US-10-858-730-83	Sequence 83, Appl	491	31	50.0	720	7	US-11-113-424-28	Sequence 28, Appl
419	31	50.0	392	6	US-10-793-626-2494	Sequence 2494, App	492	31	50.0	747	6	US-10-131-826A-426	Sequence 426, App
420	31	50.0	399	7	US-11-077-386-18	Sequence 18, Appl	493	31	50.0	747	6	US-10-973-115B-426	Sequence 426, App
421	31	50.0	401	6	US-10-949-720-419	Sequence 419, App	494	31	50.0	747	7	US-11-113-424-26	Sequence 26, Appl
422	31	50.0	404	6	US-10-793-626-2638	Sequence 2638, App	495	31	50.0	771	6	US-10-949-720-389	Sequence 389, App
423	31	50.0	414	7	US-11-115-868-2	Sequence 2, Appl	496	31	50.0	778	7	US-11-072-512-3335	Sequence 3335, App
424	31	50.0	414	7	US-11-165-305-2	Sequence 2, Appl	497	31	50.0	813	7	US-11-087-099-4670	Sequence 4670, App
425	31	50.0	419	7	US-11-126-313-25	Sequence 25, Appl	498	31	50.0	831	7	US-11-098-686-10875	Sequence 10875, A
426	31	50.0	435	7	US-11-077-386-19	Sequence 19, Appl	499	31	50.0	862	7	US-11-182-016-35	Sequence 35, Appl
427	31	50.0	437	7	US-11-087-099-11532	Sequence 11532, A	500	31	50.0	897	7	US-11-124-367A-449	Sequence 449, App
428	31	50.0	440	6	US-10-467-657-1112	Sequence 1112, App							
429	31	50.0	444	7	US-11-087-099-9103	Sequence 9103, App							
430	31	50.0	450	7	US-11-077-386-20	Sequence 20, Appl							
431	31	50.0	454	6	US-10-509-773-8	Sequence 8, Appl							
432	31	50.0	459	7	US-11-200-486-2	Sequence 2, Appl							
433	31	50.0	468	6	US-10-957-569-28	Sequence 28, Appl							
434	31	50.0	468	6	US-10-467-657-1112	Sequence 1112, App							
435	31	50.0	468	7	US-11-097-589-26	Sequence 26, Appl							
436	31	50.0	470	6	US-10-793-626-2496	Sequence 2496, App							
437	31	50.0	470	6	US-10-467-657-8420	Sequence 8420, App							
438	31	50.0	473	6	US-10-453-372-770	Sequence 770, App							
439	31	50.0	481	7	US-11-096-568A-10699	Sequence 10699, A							
440	31	50.0	482	6	US-10-467-657-7660	Sequence 7660, App							
441	31	50.0	490	7	US-11-096-568A-19018	Sequence 19018, A							
442	31	50.0	499	7	US-11-087-099-5922	Sequence 5922, App							
443	31	50.0	500	7	US-11-087-099-7042	Sequence 7042, App							
444	31	50.0	500	7	US-11-087-099-9108	Sequence 9108, App							
445	31	50.0	510	7	US-11-087-099-8796	Sequence 8796, App							
446	31	50.0	522	6	US-10-949-720-425	Sequence 425, App							
447	31	50.0	522	6	US-11-096-568A-19017	Sequence 19017, A							
448	31	50.0	524	7	US-11-096-568A-19016	Sequence 19016, A							
449	31	50.0	532	7	US-11-096-568A-2153	Sequence 2153, App							
450	31	50.0	536	7	US-11-018-868-42	Sequence 42, Appl							
451	31	50.0	537	6	US-10-949-720-424	Sequence 424, App							
452	31	50.0	552	7	US-11-096-568A-18755	Sequence 18755, A							
453	31	50.0	555	6	US-10-949-720-387	Sequence 387, App							
454	31	50.0	561	7	US-11-096-568A-2152	Sequence 2152, App							
455	31	50.0	561	7	US-11-096-568A-18754	Sequence 18754, A							
456	31	50.0	567	7	US-11-096-568A-18753	Sequence 18753, A							
457	31	50.0	570	6	US-10-949-720-386	Sequence 386, App							
458	31	50.0	570	6	US-10-949-720-412	Sequence 412, App							
459	31	50.0	577	7	US-11-200-486-6	Sequence 6, Appl							
460	31	50.0	579	6	US-10-623-155-176	Sequence 176, App							
461	31	50.0	579	6	US-10-623-155-348	Sequence 348, App							
462	31	50.0	579	6	US-10-623-155-446	Sequence 446, App							
463	31	50.0	579	6	US-10-623-155-449	Sequence 449, App							

ALIGNMENTS

RESULT 1

US-10-821-234-964

Sequence 964, Application US/10821234

Publication No. US20050255114A1

GENERAL INFORMATION:

APPLICANT: Labat, Ivan

APPLICANT: Stache-Crain, Birgit

APPLICANT: Andarmani, Susan

APPLICANT: Tang, Y. Tom

TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia

FILE REFERENCE: 821A

CURRENT APPLICATION NUMBER: US/10/821,234

CURRENT FILING DATE: 2004-04-07

PRIOR FILING DATE: 2003-04-07

NUMBER OF SEQ ID NOS: 1704

SOFTWARE: pt_seq_genes Version 1.0

SEQ ID NO 964

LENGTH: 1166

TYPE: PRT

ORGANISM: Homo sapiens

US-10-821-234-964

Query Match 77.4%; Score 48; DB 6; Length 1166;

Best Local Similarity 80.0%; Pred. No. 3.3;

Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGAGGSPG 10

Db 870 GEGGEGSPG 879

RESULT 2
US-11-186-284-33
; Sequence 33, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlögel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MP01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-33

Query Match 77.4%; Score 48; DB 7; Length 1466;
Best Local Similarity 80.0%; Pred. No. 4.2;
Matches 8; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GEGKAGSGSPG 10
|||:|||||
Db 1170 GERSESGSPG 1179

RESULT 3
US-10-514-040-4
; Sequence 4, Application US/10514040
; Publication No. US20050288223A1
; GENERAL INFORMATION:
; APPLICANT: Lucas, John
; APPLICANT: Dialynas, Deno
; TITLE OF INVENTION: OBG3 CONVERSION-DIRECTED FRAGMENTS AND OTHER COMPOSITIONS FOR TRE
; TITLE OF INVENTION: METABOLIC DISORDERS
; FILE REFERENCE: WO783
; CURRENT APPLICATION NUMBER: US/10/514,040
; CURRENT FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-514-040-4

Query Match 75.8%; Score 47; DB 6; Length 244;
Best Local Similarity 66.7%; Pred. No. 0.94;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGLL 12
||||:|||||
Db 63 GEKGEKDPGLI 74

RESULT 4

US-10-296-865-6
; Sequence 6, Application US/10296865
; Publication No. US20060035824A1
; GENERAL INFORMATION:
; APPLICANT: Lodish, Harvey
; APPLICANT: Fruebis, Joachim
; APPLICANT: Tsao, Tsu-Shuen
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: Use of ACRP30 Globular Head to Promote Increases in Muscle Mass a
; TITLE OF INVENTION: Muscle Differentiation
; FILE REFERENCE: G-087US03PCT
; CURRENT APPLICATION NUMBER: US/10/296,865
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: PCT/IB/01/01126
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/239,735
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: US 60/208,251
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 6
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-296-865-6

Query Match 75.8%; Score 47; DB 6; Length 244;
Best Local Similarity 66.7%; Pred. No. 0.94;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGLL 12
||||:|||||
Db 63 GEKGEKDPGLI 74

RESULT 5

US-11-157-947-1
; Sequence 1, Application US/11157947
; Publication No. US20050266506A1
; GENERAL INFORMATION:
; APPLICANT: TOMITA, MOTOWO
; APPLICANT: NAKANO, YASUKO
; APPLICANT: HIROSE, HIROSHI
; APPLICANT: MATSUBARA, KOICHI
; TITLE OF INVENTION: METHOD FOR DIAGNOSING OR MONITORING CARBOHYDRATE METABOLISM
; TITLE OF INVENTION: DISORDERS
; FILE REFERENCE: Q79915
; CURRENT APPLICATION NUMBER: US/11/157,947
; CURRENT FILING DATE: 2005-06-22
; PRIOR APPLICATION NUMBER: PCT/JP02/08331
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: JP 2001-248047
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 3.2
; SEQ ID NO 1
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-157-947-1

Query Match 75.8%; Score 47; DB 7; Length 244;
Best Local Similarity 66.7%; Pred. No. 0.94;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGLL 12
||||:|||||
Db 63 GEKGEKDPGLI 74

```
RESULT 6
US-11-256-802-3
; Sequence 3, Application US/11256802
; Publication No. US2006003486A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG ZACRP3
; FILE REFERENCE: 99-09
; CURRENT APPLICATION NUMBER: US/11/256,802
; CURRENT FILING DATE: 2005-10-24
; PRIOR FILING DATE: US/09/552,225
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/130,199
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-256-802-3

Query Match          75.8%; Score 47; DB 7; Length 244;
Best Local Similarity 66.7%; Pred. No. 0.94;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 12
    ||||| : |||||
Db 63 GEGKGGDPGLI 74

RESULT 7
US-11-258-647-4
; Sequence 4, Application US/11258647
; Publication No. US20060040360A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; TITLE OF INVENTION: HOMOLOG ZACRP7
; FILE REFERENCE: 99-31C2
; CURRENT APPLICATION NUMBER: US/11/258,647
; CURRENT FILING DATE: 2005-10-25
; PRIOR APPLICATION NUMBER: US 10/234,000
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US 09/577,298
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: US 60/158,448
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: US 60/145,589
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/136,289
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-258-647-4

Query Match          75.8%; Score 47; DB 7; Length 244;
Best Local Similarity 66.7%; Pred. No. 0.94;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 12
    ||||| : |||||
Db 63 GEGKGGDPGLI 74

RESULT 8
US-11-051-720-1372
; Sequence 1372, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051,720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1372
; LENGTH: 1081
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1372

Query Match          74.2%; Score 46; DB 7; Length 1081;
Best Local Similarity 72.7%; Pred. No. 6.6;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 11
    ||||| : |||||
Db 712 GEGPQKPGGL 722

RESULT 9
US-11-124-368A-329
; Sequence 329, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 329
; LENGTH: 1736
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-368A-329

Query Match          74.2%; Score 46; DB 7; Length 1736;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEGKAGSGPGL 11
    ||||| : |||||
Db 670 GEGPQKPGGL 680

RESULT 10
US-10-995-561-911
; Sequence 911, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 911
; LENGTH: 1767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-911

Query Match      74.2%; Score 46; DB 6; Length 1767;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
      |||||:||||
Db      673 GEGGPGKPG 683

RESULT 11
US-10-995-561-914
; Sequence 914, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 914
; LENGTH: 1767
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-914

Query Match      74.2%; Score 46; DB 6; Length 1767;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
      |||||:||||
Db      673 GEGGPGKPG 683

RESULT 12
US-10-995-561-912
; Sequence 912, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 912
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-912

Query Match      74.2%; Score 46; DB 6; Length 1806;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
      |||||:||||
Db      673 GEGGPGKPG 683

RESULT 13
US-10-995-561-915
; Sequence 915, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995.561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 915
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-915

Query Match      74.2%; Score 46; DB 6; Length 1806;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
      |||||:||||
Db      712 GEGGPGKPG 722

RESULT 14
US-11-051-720-1446
; Sequence 1446, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051.720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1446
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1446

Query Match      74.2%; Score 46; DB 7; Length 1806;
Best Local Similarity 72.7%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGGAGSPGL 11
      |||||:||||
Db      712 GEGGPGKPG 722

RESULT 15
US-11-051-720-1447
; Sequence 1447, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Compugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051.720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1447
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1447
```

```
Query Match      74.2%; Score 46; DB 7; Length 1806;
Best Local Similarity 72.7%; Pred. No. 11;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPL 11
      ||||| : |||||
Db      712 GEGPQKPG 722

RESULT 16
US-10-995-561-910
; Sequence 910, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 910
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-910

Query Match      74.2%; Score 46; DB 6; Length 1818;
Best Local Similarity 72.7%; Pred. No. 11;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPL 11
      ||||| : |||||
Db      724 GEGPQKPG 734

RESULT 17
US-10-995-561-913
; Sequence 913, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 913
; LENGTH: 1818
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-913

Query Match      74.2%; Score 46; DB 6; Length 1818;
Best Local Similarity 72.7%; Pred. No. 11;
Matches      8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPL 11
      ||||| : |||||
Db      724 GEGPQKPG 734

RESULT 18
US-10-501-035-331
; Sequence 331, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 331
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-331

Query Match      72.6%; Score 45; DB 6; Length 1464;
Best Local Similarity 80.0%; Pred. No. 13;
Matches      8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
      ||||| : |||||
Db      749 GPKGADGSPG 758

RESULT 19
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Query Match      72.6%; Score 45; DB 7; Length 1464;
Best Local Similarity 80.0%; Pred. No. 13;
Matches      8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
```

Query

GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-37

Query Match 69.4%; Score 43; DB 7; Length 744;
Best Local Similarity 72.7%; Pred. No. 14;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGKAGSPGL 11
|||||
DB 375 GEGPISPGI 385

RESULT 25
US-11-186-284-35
; Sequence 35, Application US/11186284
; Publication No. US2005026493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 1496
; TYPE: PRT
; ORGANISM: Homo Sapiens

US-11-186-284-35

Query Match 69.4%; Score 43; DB 7; Length 1496;
Best Local Similarity 80.0%; Pred. No. 29;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEGKAGSPG 10
|||||
DB 780 GEGKAGTAG 789

RESULT 26
US-11-181-091-6
; Sequence 6, Application US/11181091
; Publication No. US20060030046A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, Nobuto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/11/181,091
; FILING DATE: 14-Jul-2005
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999
; APPLICATION NUMBER: 08/809,156
; FILING DATE: <Unknown>
; APPLICATION NUMBER: JP 294382/1995
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: JP 051847/1996
; FILING DATE: 08-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 977.6507P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 186 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-11-181-091-6

Query Match 67.7%; Score 42; DB 7; Length 186;
Best Local Similarity 70.0%; Pred. No. 4.7;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Mon Mar 13 10:53:09 2006

```
QY 1 GEGGAEGSPG 10
    ||||| |:||
Db 97 GEGGGRGTGP 106

RESULT 27
US-11-174-150-30
; Sequence 30, Application US/11174150
; Publication No. US20050260714A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhacaying
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50022
; CURRENT APPLICATION NUMBER: US/11/174,150
; CURRENT FILING DATE: 2005-07-01
; PRIOR APPLICATION NUMBER: US/10/257,174
; PRIOR FILING DATE: 2002-10-10
; PRIOR APPLICATION NUMBER: PCT/US01/11797
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/196,603
; PRIOR FILING DATE: 2000-04-13
; PRIOR APPLICATION NUMBER: 60/199,417
; PRIOR FILING DATE: 2000-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-174-150-30

Query Match 67.7%; Score 42; DB 7; Length 287;
Best Local Similarity 72.7%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GEGGAEGSPGL 11
    ||||| |:||
Db 106 GEGGDSGRFGL 116

RESULT 28
US-11-181-091-7
; Sequence 7, Application US/11181091
; Publication No. US20060030046A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, Nobuto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/11/181,091
; FILING DATE: 14-Jul-2005

QY 1 GEGGAEGSPG 10
    ||||| |:||
Db 375 GEGGGRGTGP 384

RESULT 29
US-11-181-091-8
; Sequence 8, Application US/11181091
; Publication No. US20060030046A1
; GENERAL INFORMATION:
; APPLICANT: Asada, Kiyozo
; Uemori, Takashi
; Ueno, Takashi
; Koyama, Nobuto
; Hashino, Kimikazu
; Kato, Ikunoshin
; TITLE OF INVENTION: METHOD FOR GENE TRANSFER INTO TARGET
; CELLS WITH RETROVIRUS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/11/181,091
; FILING DATE: 14-Jul-2005
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/775,964
; FILING DATE: 20-Feb-2001
; APPLICATION NUMBER: US/09/366,009
; FILING DATE: 02-Aug-1999

CLASSIFICATION: <Unknown>
PRIOR APPLICATION NUMBER: US/09/775,964
FILING DATE: 20-Feb-2001
APPLICATION NUMBER: US/09/366,009
FILING DATE: 02-Aug-1999
APPLICATION NUMBER: 08/809,156
FILING DATE: 02-Aug-1999
APPLICATION NUMBER: JP 294382/1995
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: JP 051847/1996
FILING DATE: 08-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 977.6507P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 464 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-11-181-091-7

Query Match 67.7%; Score 42; DB 7; Length 464;
Best Local Similarity 70.0%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
```



```

/ APPLICATION NUMBER: 08/809,156
/ FILING DATE: <Unknown>
/ APPLICATION NUMBER: JP 294382/1995
/ FILING DATE: 13-NOV-1995
/ APPLICATION NUMBER: JP 051847/1996
/ FILING DATE: 08-MAR-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weiser, Gerard J.
/ REGISTRATION NUMBER: 19,763
/ REFERENCE/DOCKET NUMBER: 977.6507P
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 215-875-8383
/ TELEFAX: 215-875-8394
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 489 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: <Unknown>
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-11-181-091-8
Query Match 67.7%; Score 42; DB 7; Length 489;
Best Local Similarity 70.0%; Pred.No.13;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEKGAGSGPG 10
||||| |||
DB 375 GEGGGRGTPG 384

RESULT 30
US-10-330-773-700
; Sequence 700, Application US/10330773
; Publication No. US20060040262A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001300
; CURRENT APPLICATION NUMBER: US/10/330,773
; CURRENT FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 981
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 700
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-330-773-700
Query Match 67.7%; Score 42; DB 6; Length 822;
Best Local Similarity 70.0%; Pred.No.23;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GEKGAGSGPG 10
|:|:|:|:|
DB 60 GSRGAKGSPG 69

RESULT 31
US-10-821-234-1431
; Sequence 1431, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07

```

```

; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821.234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1182
; LENGTH: 1874
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1182

Query Match      67.7%; Score 42; DB 6; Length 1874;
Best Local Similarity 70.0%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPG 10
Db 808 GEKGQGP 817

RESULT 34
US-11-052-554A-368
; Sequence 368, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 368
; LENGTH: 2551
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae R6
US-11-052-554A-368

Query Match      67.7%; Score 42; DB 7; Length 2551;
Best Local Similarity 72.7%; Pred. No. 74;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGL 11
Db 688 GEKGAGSPGL 698

RESULT 35
US-10-514-057-1
; Sequence 1, Application US/10514057
; Publication No. US20050255547A1
; GENERAL INFORMATION:
; APPLICANT: Tschopp, Jurq
; TITLE OF INVENTION: Novel hexamers of receptors, members of the TNF-receptor family,
; TITLE OF INVENTION: their use in therapy and pharmaceutical compositions comprising
; TITLE OF INVENTION: the same
; FILE REFERENCE: 11436*15
; CURRENT APPLICATION NUMBER: US/10/514,057
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: PCT/EP02/12186
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: PCT/EP02/05103
; PRIOR FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.2

```

```

; SEQ ID NO 1
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-514-057-1

Query Match      66.1%; Score 41; DB 6; Length 66;
Best Local Similarity 66.7%; Pred. No. 2.3;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GEKGAGSPGLL 12
Db 22 GEKGAGDAGLL 33

RESULT 36
US-10-131-826A-362
; Sequence 362, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 362
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-362

Query Match      66.1%; Score 41; DB 6; Length 243;
Best Local Similarity 72.7%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```
Qy      1 GEGGAGSPGL 11
Db      63 GEGGGRPGL 73

RESULT 37
US-10-973-115B-362
; Sequence 362, Application US/10973115B
; Publication No. US20060040351A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 39870-3330RIC300C1
; CURRENT APPLICATION NUMBER: US/10/973,115B
; PRIOR FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 10/145,747
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: US 10/028,072
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US 09/581,742
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: PCT/US00/05746
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/135,736
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: US 60/123,090
; PRIOR FILING DATE: 1999-03-05
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 362
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-973-115B-362

Query Match      66.1%; Score 41; DB 6; Length 243;
Best Local Similarity 72.7%; Pred. No. 9.2;
Matches      8; Conservative      0; Mismatches      3; Indels      0; Gaps      0;

Qy      1 GEGGAGSPGL 11
Db      63 GEGGGRPGL 73

RESULT 38
US-10-514-040-2
; Sequence 2, Application US/10514040
; Publication No. US20050288223A1
; GENERAL INFORMATION:
; APPLICANT: Lucas, John
; APPLICANT: Dialynas, Deno
; TITLE OF INVENTION: ORG3 CONVERSION-DIRECTED FRAGMENTS AND OTHER COMPOSITIONS FOR TRE
; TITLE OF INVENTION: METABOLIC DISORDERS
; FILE REFERENCE: WO783
```

```
; CURRENT APPLICATION NUMBER: US/10/514,040
; CURRENT FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-514-040-2

Query Match      66.1%; Score 41; DB 6; Length 247;
Best Local Similarity 66.7%; Pred. No. 9.3;
Matches      8; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

Qy      1 GEGGAGSPGL 12
Db      66 GEGGKDAGLL 77

RESULT 39
US-10-296-865-2
; Sequence 2, Application US/10296865
; Publication No. US20060035824A1
; GENERAL INFORMATION:
; APPLICANT: Lodish, Harvey
; APPLICANT: Truebis, Joachim
; APPLICANT: Tsao, Tsu-Shuen
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: Use of ACRP30 Globular Head to Promote Increases in Muscle Mass at
; FILE REFERENCE: G-087US03PCT
; CURRENT APPLICATION NUMBER: US/10/296,865
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: PCT/IB/01/01126
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/239,735
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: US 60/208,251
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO 2
; LENGTH: 247
; TYPE: PRT
; ORGANISM: mus musculus
US-10-296-865-2

Query Match      66.1%; Score 41; DB 6; Length 247;
Best Local Similarity 66.7%; Pred. No. 9.3;
Matches      8; Conservative      1; Mismatches      3; Indels      0; Gaps      0;

Qy      1 GEGGAGSPGL 12
Db      66 GEGGKDAGLL 77

RESULT 40
US-10-296-865-4
; Sequence 4, Application US/10296865
; Publication No. US20060035824A1
; GENERAL INFORMATION:
; APPLICANT: Lodish, Harvey
; APPLICANT: Truebis, Joachim
; APPLICANT: Tsao, Tsu-Shuen
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: Use of ACRP30 Globular Head to Promote Increases in Muscle Mass at
; FILE REFERENCE: G-087US03PCT
; CURRENT APPLICATION NUMBER: US/10/296,865
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: PCT/IB/01/01126
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/239,735
```

```
; PRIOR FILING DATE: 2000-10-11
; PRIOR APPLICATION NUMBER: US 60/208,251
; PRIOR FILING DATE: 2000-05-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent.pm
; SEQ ID NO. 4
; LENGTH: 247
; TYPE: PRT
; ORGANISM: mus musculus
US-10-296-865-4

Query Match      66.1%; Score 41; DB 6; Length 247;
Best Local Similarity 66.7%; Pred. No. 9.3;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 GEGKAGSGPGLL 12
        ||||| : |||
DB      66 GEGKKGKGDAGLL 77

RESULT 41
US-10-514-057-6
; Sequence 6, Application US/10514057
; Publication No. US20050255547A1
; GENERAL INFORMATION:
; APPLICANT: Techopp, Jurq
; APPLICANT: Schneider, Pascal
; TITLE OF INVENTION: Novel hexamers of receptors, members of the TNF-receptor family,
; TITLE OF INVENTION: their use in therapy and pharmaceutical compositions comprising
; TITLE OF INVENTION: the same
; FILE REFERENCE: 11436*15
; CURRENT APPLICATION NUMBER: US/10/514,057
; CURRENT FILING DATE: 2004-11-08
; PRIOR APPLICATION NUMBER: PCT/EP02/12186
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: PCT/EP02/05103
; PRIOR FILING DATE: 2002-05-08
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 6
; LENGTH: 334
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Synthetic Construct
US-10-514-057-6

Query Match      66.1%; Score 41; DB 6; Length 334;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 GEGKAGSGPGLL 12
        ||||| : |||
DB      263 GEGKKGKGDAGLL 274

RESULT 42
US-10-995-561-983
; Sequence 983, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 983
; LENGTH: 828
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-995-561-983

Query Match      66.1%; Score 41; DB 6; Length 828;
Best Local Similarity 54.5%; Pred. No. 33;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 GEGKAGSGPGL 11
        |||| : |||
DB      280 GQKGRQGDPGI 290

RESULT 43
US-10-995-561-981
; Sequence 981, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 981
; LENGTH: 918
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-981

Query Match      66.1%; Score 41; DB 6; Length 918;
Best Local Similarity 54.5%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 GEGKAGSGPGL 11
        |||| : |||
DB      280 GQKGRQGDPGI 290

RESULT 44
US-10-995-561-982
; Sequence 982, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 982
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-982

Query Match      66.1%; Score 41; DB 6; Length 1019;
Best Local Similarity 54.5%; Pred. No. 41;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 GEGKAGSGPGL 11
        |||| : |||
DB      280 GQKGRQGDPGI 290

RESULT 45
US-10-821-234-914
; Sequence 914, Application US/10821234
```

```
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 914
; LENGTH: 1532
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-914

Query Match      66.1%; Score 41; DB 6; Length 1532;
Best Local Similarity 70.0%; Pred. No. 63;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
      |||:|||
Db      679 GERGAGGPG 688

RESULT 46
US-11-245-689-17
; Sequence 17, Application US/11245689
; Publication No. US2006003536A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 17
; LENGTH: 177
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc11 protein
US-11-245-689-17

Query Match      64.5%; Score 40; DB 7; Length 177;
Best Local Similarity 60.0%; Pred. No. 9.6;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
      |||:|||
Db      31 GERGQGNPG 40

RESULT 47
US-11-245-689-18
; Sequence 18, Application US/11245689
; Publication No. US2006003536A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
```

```
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 18
; LENGTH: 180
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc11 protein
US-11-245-689-18

Query Match      64.5%; Score 40; DB 7; Length 180;
Best Local Similarity 60.0%; Pred. No. 9.8;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
      |||:|||
Db      34 GERGQGNPG 43

RESULT 48
US-11-245-689-15
; Sequence 15, Application US/11245689
; Publication No. US2006003536A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 15
; LENGTH: 186
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc11 protein
US-11-245-689-15

Query Match      64.5%; Score 40; DB 7; Length 186;
Best Local Similarity 70.0%; Pred. No. 10;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEGKAGSGPG 10
      |||:|||
Db      49 GERGAGTGP 58

RESULT 49
US-11-135-855-30
; Sequence 30, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
```

; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-30

Query Match 64.5%; Score 40; DB 7; Length 288;
Best Local Similarity 54.5%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 1 GEGGAGSPGL 11
|:|:|:|:|
DB 124 GDRGQDPGL 134

RESULT 50

US-11-135-855-31
; Sequence 31, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; CURRENT FILING DATE: 2005-03-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 31
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-31

Query Match 64.5%; Score 40; DB 7; Length 303;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 1 GEGGAGSPGL 11
|:|:|:|:|
DB 139 GDRGQDPGL 149

RESULT 51

US-11-258-647-2
; Sequence 2, Application US/11258647
; Publication No. US2006040360A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
; FILE REFERENCE: 99-31C2
; CURRENT APPLICATION NUMBER: US/11/258,647
; CURRENT FILING DATE: 2005-10-25
; PRIOR APPLICATION NUMBER: US 10/234,000

; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US 09/577,298
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: US 60/158,448
; PRIOR FILING DATE: 1999-10-07
; PRIOR APPLICATION NUMBER: US 60/145,589
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/136,289
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 303
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-258-647-2

Query Match 64.5%; Score 40; DB 7; Length 303;
Best Local Similarity 54.5%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 2; Indels 2; Gaps 0;

QY 1 GEGGAGSPGL 11
|:|:|:|:|
DB 139 GDRGQDPGL 149

RESULT 52

US-11-245-689-44
; Sequence 44, Application US/11245689
; Publication No. US2006003336A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 44
; LENGTH: 313
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: p176
US-11-245-689-44

Query Match 64.5%; Score 40; DB 7; Length 313;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GEGGAGSPG 10
|:|:|:|:|
DB 133 GEGGAGTPG 142

RESULT 53

US-10-055-877-289
; Sequence 289, Application US/10055877
; Publication No. US20050288241A1
; GENERAL INFORMATION:
; APPLICANT: DeCristofaro, Marc
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Miller, Charles
; APPLICANT: Tchernev, Velizar
; APPLICANT: Zhong, Mei
; APPLICANT: Anderson, David
; APPLICANT: Ballinger, Robert

RESULT 55
US-10-821-234-1528
; Sequence 1528, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan

; PRIOR FILING DATE: 2001-03-14
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ;

```

; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821.234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1528
; LENGTH: 629
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1528

Query Match      64.5%; Score 40; DB 6; Length 629;
Best Local Similarity 58.3%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GEKGAGSGPGLL 12
Db      20 GQGTAGAPGLL 31

RESULT 56
US-11-186-284-39
; Sequence 39, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MP001-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 744
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-39

Query Match      64.5%; Score 40; DB 7; Length 744;
Best Local Similarity 63.6%; Pred. No. 43;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEKGAGSGPGL 11
Db      375 GEKGPIGAGGI 385

RESULT 57
US-11-052-554A-148
; Sequence 148, Application US/11052554A
```

```

; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 148
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-148

Query Match      64.5%; Score 40; DB 7; Length 749;
Best Local Similarity 88.9%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      4 GAEGSPGLL 12
Db      740 GAEGLPGLL 748

RESULT 58
US-11-245-689-19
; Sequence 19, Application US/11245689
; Publication No. US2006003536A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 19
; LENGTH: 75
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc11 protein
US-11-245-689-19

Query Match      62.9%; Score 39; DB 7; Length 75;
Best Local Similarity 70.0%; Pred. No. 5.7;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEKGAGSPG 10
Db      1 GEKGDPGAPG 10

RESULT 59
US-11-256-802-4
; Sequence 4, Application US/11256802
; Publication No. US20060034866A1
; GENERAL INFORMATION:
; APPLICANT: Piddington, Christopher S.
; APPLICANT: Bishop, Paul
; TITLE OF INVENTION: ADIPOCYTE-SPECIFIC PROTEIN HOMOLOG ZACRP3
; FILE REFERENCE: 99-09
```



```
; CURRENT APPLICATION NUMBER: US/11/256,802
; CURRENT FILING DATE: 2005-10-24
; PRIOR APPLICATION NUMBER: US/09/552,225
; PRIOR FILING DATE: 2000-04-19
; PRIOR APPLICATION NUMBER: 60/130,199
; PRIOR FILING DATE: 1999-04-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-256-802-4

Query Match      62.9%; Score 39; DB 7; Length 245;
Best Local Similarity 63.6%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GEKGAEGSPGL 11
Db      70 GPKGQGEKPG 80

RESULT 60
US-11-051-720-1334
; Sequence 1334, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Comugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051,720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1334
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1334

Query Match      62.9%; Score 39; DB 7; Length 258;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GEKGAEGSPG 10
Db      97 GEKGEKPG 106

RESULT 61
US-11-051-720-1434
; Sequence 1434, Application US/11051720
; Publication No. US20060046257A1
; GENERAL INFORMATION:
; APPLICANT: Comugen Ltd
; TITLE OF INVENTION: NOVEL NUCLEOTIDE AND AMINO ACID SEQUENCES, AND ASSAYS AND METHODS
; TITLE OF INVENTION: THEREOF FOR DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 1847.1002
; CURRENT APPLICATION NUMBER: US/11/051,720
; CURRENT FILING DATE: 2005-01-27
; NUMBER OF SEQ ID NOS: 1780
; SEQ ID NO 1434
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-051-720-1434

Query Match      62.9%; Score 39; DB 7; Length 258;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 GEKGAEGSPG 10
```

```
Db      97 GEKGEKPG 106

RESULT 62
US-11-122-524-2
; Sequence 2, Application US/11122524
; Publication No. US20060003355A1
; GENERAL INFORMATION:
; APPLICANT: BAKER, KEVIN P.
; APPLICANT: CHEN, JIAN
; APPLICANT: DESNOYERS, IJC
; APPLICANT: GODDARD, AUDREY
; APPLICANT: GODOWSKI, PAUL J.
; APPLICANT: GURNEY, AUSTIN L.
; APPLICANT: PAN, JAMES
; APPLICANT: SMITH, VICTORIA
; APPLICANT: WATANABE, COLIN K.
; APPLICANT: WOOD, WILLIAM I.
; APPLICANT: ZHANG, ZEMIN
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES
; TITLE OF INVENTION: AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: 39780-3430RIC377C1
; CURRENT APPLICATION NUMBER: US/11/122,524
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: US 10/199,461
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US 10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/380,137
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 60/096,687
; PRIOR FILING DATE: 1998-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-122-524-2

Query Match      62.9%; Score 39; DB 7; Length 259;
Best Local Similarity 70.0%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GEKGAEGSPG 10
Db      108 GDKGEMGSPG 117

RESULT 63
US-11-113-424-43
; Sequence 43, Application US/1113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gargolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
```

;; PRIOR FILING DATE: 2001-08-29
;; PRIOR APPLICATION NUMBER: 60/307,506
;; PRIOR FILING DATE: 2001-07-24
;; PRIOR APPLICATION NUMBER: 60/322,358
;; PRIOR FILING DATE: 2001-09-14
;; PRIOR APPLICATION NUMBER: 60/294,075
;; PRIOR FILING DATE: 2001-05-29
;; PRIOR APPLICATION NUMBER: 60/288,153
;; PRIOR FILING DATE: 2001-05-02
;; NUMBER OF SEQ ID NOS: 190
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 43
;; LENGTH: 278
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-11-113-424-43

Query Match 62.9%; Score 39; DB 7; Length 278;
Best Local Similarity 70.0%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GKGAGSGSPG 10
Db 127 GDKGMSGSPG 136

RESULT 64
US-11-258-647-5
;; Sequence 5, Application US/11258647
;; Publication No. US20060040360A1
;; GENERAL INFORMATION:
;; APPLICANT: Piddington, Christopher S.
;; APPLICANT: Sheppard, Paul O.
;; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN
;; FILE OF INVENTION: HOMOLOG ZACRP7
;; CURRENT APPLICATION NUMBER: US/11/258,647
;; CURRENT FILING DATE: 2005-10-25
;; PRIOR APPLICATION NUMBER: US 10/234,000
;; PRIOR FILING DATE: 2002-08-30
;; PRIOR APPLICATION NUMBER: US 09/577,298
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: US 60/158,448
;; PRIOR FILING DATE: 1999-10-07
;; PRIOR APPLICATION NUMBER: US 60/145,589
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/136,289
;; PRIOR FILING DATE: 1999-05-27
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5
;; LENGTH: 285
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-11-258-647-5

Query Match 62.9%; Score 39; DB 7; Length 285;
Best Local Similarity 63.6%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 GKGAGSGPGL 11
Db 127 GPKGKGEPGL 137

RESULT 65
US-11-258-647-15
;; Sequence 15, Application US/11258647
;; Publication No. US20060040360A1
;; GENERAL INFORMATION:
;; APPLICANT: Piddington, Christopher S.
;; APPLICANT: Sheppard, Paul O.
;; TITLE OF INVENTION: ADIPOCYTE COMPLEMENT RELATED PROTEIN

;; TITLE OF INVENTION: HOMOLOG ZACRP7
;; FILE REFERENCE: 99-31C2
;; CURRENT APPLICATION NUMBER: US/11/258,647
;; CURRENT FILING DATE: 2005-10-25
;; PRIOR APPLICATION NUMBER: US 10/234,000
;; PRIOR FILING DATE: 2002-08-30
;; PRIOR APPLICATION NUMBER: US 09/577,298
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: US 60/158,448
;; PRIOR FILING DATE: 1999-10-07
;; PRIOR APPLICATION NUMBER: US 60/145,589
;; PRIOR FILING DATE: 1999-07-26
;; PRIOR APPLICATION NUMBER: US 60/136,289
;; PRIOR FILING DATE: 1999-05-27
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 15
;; LENGTH: 289
;; TYPE: PRT
;; ORGANISM: Mus musculus
US-11-258-647-15

Query Match 62.9%; Score 39; DB 7; Length 289;
Best Local Similarity 54.5%; Pred. No. 24;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GKGAGSGPGL 11
Db 125 GDRGDQDGPGL 135

RESULT 66
US-11-072-512-3814
;; Sequence 3814, Application US/11072512
;; Publication No. US20060029945A1
;; GENERAL INFORMATION:
;; APPLICANT: ISOGAI, TAKAO
;; APPLICANT: SUGIYAMA, TOMOYASU
;; APPLICANT: OTSUKI, TETSUJI
;; APPLICANT: WAKAMATSU, AI
;; APPLICANT: SATO, HIROYUKI
;; APPLICANT: ISHII, SHIZUKO
;; APPLICANT: YAMAMOTO, JUN-ICHI
;; APPLICANT: ISONO, YUUKO
;; APPLICANT: HIO, YURI
;; APPLICANT: OTSUKA, KAORU
;; APPLICANT: NAGAI, KEIICHI
;; APPLICANT: IRIE, RYOTARO
;; APPLICANT: TAMECHIKA, ICHIRO
;; APPLICANT: SEKI, NAOHICO
;; APPLICANT: YOSHIKAWA, TSUTOMU
;; APPLICANT: OTSUKA, MOTYUKI
;; APPLICANT: NAGAHARI, KENJI
;; APPLICANT: MASUHO, YASUHIKO
;; TITLE OF INVENTION: Novel full length cDNA
;; FILE REFERENCE: 084335-0191
;; CURRENT APPLICATION NUMBER: US/11/072,512
;; CURRENT FILING DATE: 2005-03-07
;; PRIOR APPLICATION NUMBER: US 60/350,978
;; PRIOR FILING DATE: 2002-01-25
;; PRIOR APPLICATION NUMBER: JP 2001-379298
;; PRIOR FILING DATE: 2001-11-05
;; NUMBER OF SEQ ID NOS: 4096
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 3814
;; LENGTH: 571
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-11-072-512-3814

Query Match 62.9%; Score 39; DB 7; Length 571;
Best Local Similarity 76.9%; Pred. No. 48;
Matches 10; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy 1 GEK--GAEQSPGL 11
| | | | | | | | | |
Db 6 GEKRGAGSGPKL 18
| | | | | | | | | |
RESULT 67
US-11-169-041-180
; Sequence 180, Application US/11169041
; Publication No. US20060019284A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES FOR PREDICTING ACTIVITY OF
; TITLE OF INVENTION: COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE
; TITLE OF INVENTION: KINASES AND/OR PROTEIN TYROSINE KINASE PATHWAYS IN LUNG CANCER
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 10001 NP
; CURRENT APPLICATION NUMBER: US/11/169,041
; CURRENT FILING DATE: 2005-06-28
; PRIOR APPLICATION NUMBER: 60/584,405
; PRIOR FILING DATE: 2004-06-30
; NUMBER OF SEQ ID NOS: 527
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 180
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-169-041-180
Query Match 62.9%; Score 39; DB 7; Length 1028;
Best Local Similarity 70.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GEKGAEGSPG 10
| | | | | | | | | |
Db 278 GEKGAEGDPG 287
| | | | | | | | | |
RESULT 68
US-10-220-824-8
; Sequence 8, Application US/10220824
; Publication No. US20050277603A1
; GENERAL INFORMATION:
; APPLICANT: Viomed Limited
; TITLE OF INVENTION: Compositions for gene therapy of rheumatoid arthritis including a
; TITLE OF INVENTION: Gene encoding an anti-angiogenic protein or parts thereof
; FILE REFERENCE: OP0208/PCT
; CURRENT APPLICATION NUMBER: US/10/220,824
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: KR 2001-0000691
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 8
; LENGTH: 1516
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-220-824-8
Query Match 62.9%; Score 39; DB 6; Length 1516;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GEKGAEGSPG 10
| | | | | | | | | |
Db 745 CHKSGKAGP 754
| | | | | | | | | |
RESULT 69
US-11-096-070-34
; Sequence 34, Application US/11096070
; Publication No. US20050287098A1
; GENERAL INFORMATION:

; APPLICANT: SUN, TUNG-TIEN
; APPLICANT: CAO, QIONG
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR CONTROLLING HAIR GROWTH
; FILE REFERENCE: 71369,274US2
; CURRENT APPLICATION NUMBER: US/11/096,070
; CURRENT FILING DATE: 2005-03-31
; PRIOR APPLICATION NUMBER: 60/558,341
; PRIOR FILING DATE: 2004-03-31
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 3.3
; SEQ ID NO 34
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Rattus sp.
US-11-096-070-34
Query Match 61.3%; Score 38; DB 7; Length 117;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GEKGAEGSPGL 11
| | | | | | | | | |
Db 19 GHNGSDGQPL 29
| | | | | | | | | |
RESULT 70
US-11-245-689-32
; Sequence 32, Application US/11245689
; Publication No. US2006003536A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 32
; LENGTH: 117
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Scl2 protein
US-11-245-689-32
Query Match 61.3%; Score 38; DB 7; Length 117;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GEKGAEGSPG 10
| | | | | | | | | |
Db 22 GERKEGEPG 31
| | | | | | | | | |
RESULT 71
US-11-113-424-44
; Sequence 44, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590

; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 44
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-44

Query Match 61.3%; Score 38; DB 7; Length 199;
Best Local Similarity 60.0%; Pred. No. 23;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
|:|:|:|:
Db 47 GQKSGMGAP 56

RESULT 72
US-11-245-689-36
; Sequence 36, Application US/11245689
; Publication No. US20060035336A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 36
; LENGTH: 234
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc12 protein
US-11-245-689-36

Query Match 61.3%; Score 38; DB 7; Length 234;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
|:|:|:|:
Db 109 GERKEGEPG 118

RESULT 73
US-11-245-689-37
; Sequence 37, Application US/11245689
; Publication No. US20060035336A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins

; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 37
; LENGTH: 237
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc12 protein
US-11-245-689-37

Query Match 61.3%; Score 38; DB 7; Length 237;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
|:|:|:|:
Db 112 GERKEGEPG 121

RESULT 74
US-11-245-689-41
; Sequence 41, Application US/11245689
; Publication No. US20060035336A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 41
; LENGTH: 237
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: recombinant Sc12 protein
US-11-245-689-41

Query Match 61.3%; Score 38; DB 7; Length 237;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GKGAGSGPG 10
|:|:|:|:
Db 103 GERKEGEPG 112

RESULT 75
US-11-245-689-43
; Sequence 43, Application US/11245689
; Publication No. US20060035336A1
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Lukomski, Slawomir
; APPLICANT: Xu, Yi
; TITLE OF INVENTION: Prokaryotic Collagen-Like Proteins

```
; TITLE OF INVENTION: And Uses Thereof
; FILE REFERENCE: D6562
; CURRENT APPLICATION NUMBER: US/11/245,689
; CURRENT FILING DATE: 2005-10-07
; PRIOR APPLICATION NUMBER: US/10/830,792
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/464,816
; PRIOR FILING DATE: 2003-04-23
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 43
; LENGTH: 237
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: collagen-like region of p163
US-11-245-689-43

Query Match      61.3%; Score 38; DB 7; Length 237;
Best Local Similarity 60.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GKGAGGSPG 10
      ||:|:|
Db      112 GERGEKGEPG 121
```

Search completed: March 11, 2006, 12:12:02
Job time : 18.8 secs

WEST Search History

DATE: Monday, March 13, 2006

Hide?	<u>Set Name</u>	<u>Query</u>	<u>Hit Count</u>
		<i>DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI; THES=ASSIGNEE; PLUR=YES; OP=ADJ</i>	
<input type="checkbox"/>	L2	5788966.pn.	2
<input type="checkbox"/>	L1	9961040.pn.	3

END OF SEARCH HISTORY

This Page Blank (uspto)